



*2nd edition*

ENCYCLOPEDIA OF  
**PHILOSOPHY**



PRICE – SEXTUS EMPIRICUS

8  
volume

*2nd edition*

ENCYCLOPEDIA OF  
PHILOSOPHY

DONALD M. BORCHERT

*Editor in Chief*

**MACMILLAN REFERENCE USA**

*An imprint of Thomson Gale, a part of The Thomson Corporation*

**THOMSON**  
—★—™  
**GALE**



## Encyclopedia of Philosophy, Second Edition

Donald M. Borchert, Editor in Chief

© 2006 Thomson Gale, a part of the Thomson Corporation.

Thomson, Star Logo and Macmillan Reference USA are trademarks and Gale is a registered trademark used herein under license.

For more information, contact  
Macmillan Reference USA  
An imprint of Thomson Gale  
27500 Drake Rd.  
Farmington Hills, MI 48331-3535  
Or you can visit our internet site at  
<http://www.gale.com>

### ALL RIGHTS RESERVED

No part of this work covered by the copyright hereon may be reproduced or used in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, Web distribution, or information storage retrieval systems—without the written permission of the publisher.

For permission to use material from this product, submit your request via Web at <http://www.gale-edit.com/permissions>, or you may download our Permissions Request form and submit your request by fax or mail to:

*Permissions*  
Thomson Gale  
27500 Drake Rd.  
Farmington Hills, MI 48331-3535  
Permissions Hotline:  
248-699-8006 or 800-877-4253 ext. 8006  
Fax: 248-699-8074 or 800-762-4058

Since this page cannot legibly accommodate all copyright notices, the acknowledgments constitute an extension of the copyright notice.

While every effort has been made to ensure the reliability of the information presented in this publication, Thomson Gale does not guarantee the accuracy of the data contained herein. Thomson Gale accepts no payment for listing; and inclusion in the publication of any organization, agency, institution, publication, service, or individual does not imply endorsement of the editors or publisher. Errors brought to the attention of the publisher and verified to the satisfaction of the publisher will be corrected in future editions.

### LIBRARY OF CONGRESS CATALOGING-IN-PUBLICATION DATA

Encyclopedia of philosophy / Donald M. Borchert, editor in chief.—2nd ed.  
p. cm.

Includes bibliographical references and index.  
ISBN 0-02-865780-2 (set hardcover : alk. paper)—  
ISBN 0-02-865781-0 (vol 1)—ISBN 0-02-865782-9 (vol 2)—  
ISBN 0-02-865783-7 (vol 3)—ISBN 0-02-865784-5 (vol 4)—  
ISBN 0-02-865785-3 (vol 5)—ISBN 0-02-865786-1 (vol 6)—  
ISBN 0-02-865787-X (vol 7)—ISBN 0-02-865788-8 (vol 8)—  
ISBN 0-02-865789-6 (vol 9)—ISBN 0-02-865790-X (vol 10)  
1. Philosophy—Encyclopedias. I. Borchert, Donald M., 1934-

B51.E53 2005  
103—dc22

2005018573

This title is also available as an e-book.

ISBN 0-02-866072-2

Contact your Thomson Gale representative for ordering information.

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1



*contents*

***volume 1***

PREFACE TO 2ND EDITION  
INTRODUCTION TO 1ST EDITION  
LIST OF CONTRIBUTORS  
LIST OF ARTICLES

ENCYCLOPEDIA OF  
**PHILOSOPHY**  
*2nd edition*

Abbagnano–Byzantine Philosophy

***volume 2***

Cabanis–Destutt de Tracy

***volume 3***

Determinables–Fuzzy Logic

***volume 4***

Gadamer–Just War Theory

***volume 5***

Kabbalah–Marxist Philosophy

***volume 6***

Masaryk–Nussbaum

***volume 7***

Oakeshott–Presupposition

***volume 8***

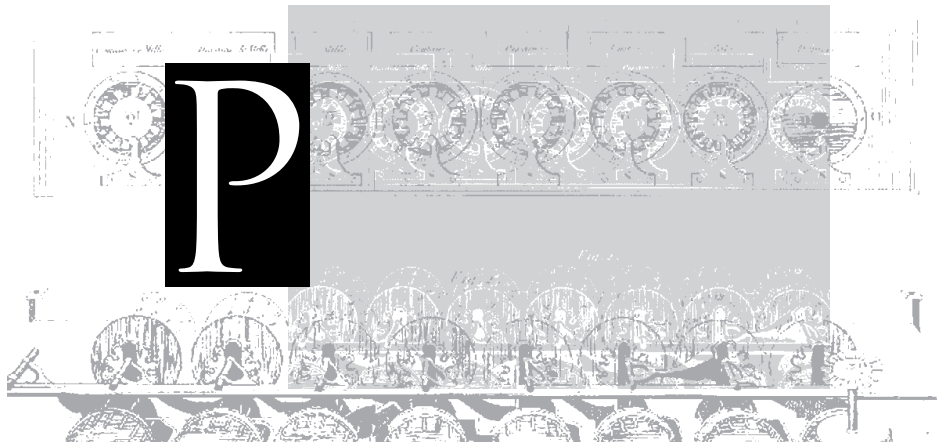
Price–Sextus Empiricus

***volume 9***

Shaftesbury–Zubiri

***volume 10***

APPENDIX: ADDITIONAL ARTICLES  
THEMATIC OUTLINE  
BIBLIOGRAPHIES  
INDEX



## PRICE, RICHARD

(1723–1791)

Richard Price, a Welsh dissenting preacher, moral philosopher, and actuary, was born at Tynon, Llangeinor, Glamorganshire. His father, Rees, was a dissenting minister with extreme Calvinist opinions. Richard Price was educated at a number of different academies, finally entering Coward's Academy in London, where he remained for the years 1740–1744. He was ordained at the age of twenty-one and began his ministerial career as a domestic chaplain. He later served a number of London congregations, notably those at Stoke Newington, where he lived, and at the Gravel-Pit Meeting House in Hackney. Price was buried in the cemetery at Bunhill Fields; his friend Joseph Priestley preached the funeral oration.

In addition to his writings on moral philosophy, Price wrote with considerable influence on financial and political questions. His papers on life expectancy and on calculating the values of reversionary payments were instrumental in reforming the actuarial basis of the insurance and benefit societies of the time. His paper on the public debt is said to have led William Pitt, the prime minister, to reestablish the sinking fund to extinguish England's national debt. In his pamphlet *Observations on*

*the Nature of Civil Liberty, the Principles of Government, and the Justice and Policy of the War with America* (London, February 8, 1776), Price defended the American cause. The widespread circulation and generally favorable acceptance of this work is said to have encouraged the American decision for a declaration of independence. Price had become friendly with Benjamin Franklin during the latter's stay in London, and in 1778 the Continental Congress moved to grant Price American citizenship if he would come to America and serve as an adviser on the management of American finances. He was grateful for the invitation but did not accept it. Price also regarded the French Revolution with approval, which he expressed, along with an appeal for reform in England, in his *Discourse on the Love of Our Country* (1789). Edmund Burke's *Reflections on the Revolution in France* (1790) was written in reply.

Price is also the author of *Four Dissertations*: I. "On Providence"; II. "On Prayer"; III. "On the Reasons for expecting that virtuous Men shall meet after death in a State of Happiness"; IV. "On the Importance of Christianity, the Nature of Historical Evidence, and Miracles" (London, 1767). In the fourth of these dissertations Price criticized David Hume's "Of Miracles." Hume was grateful for the civility with which Price argued, and he wrote

to Price that the light in which he put this controversy was “new and plausible and ingenious, and perhaps solid. But I must have some more time to weigh it, before I can pronounce this judgment with satisfaction to myself.”

## MORAL PHILOSOPHY

Price’s contribution to moral philosophy is *A Review of the Principal Questions in Morals* (London, 1758; corrected editions in 1769 and 1787). Price criticized the moral-sense doctrines of Francis Hutcheson in order to clear them away and make room for an account of immutable right and wrong, derived from Samuel Clarke.

Price says that we may have three different perceptions concerning the actions of moral agents. We may notice whether they are right or wrong, whether they are beautiful or ugly, and whether they are of good or ill desert. By talking of perceptions here, he shows that he has accepted the premise, of Lockean origin, that all knowledge is to be accounted for as some kind of perception by one of our faculties. Thus, Price’s first question, “How do we know right?” is treated as a search for the faculty by means of which we obtain our ideas of right and wrong. He considers Hutcheson’s answer that our moral ideas come to us by the way of a moral sense, and he understands Hutcheson to be claiming that this sense is “a power within us, different from reason; which renders certain actions pleasing and others displeasing to us.”

Price objects to this doctrine because of certain consequences that he believes are implied by it. Our approval and disapproval of actions appear to depend on the way our minds work or, to carry the matter back a step, on the way God has made them to work. Thus, our judgments of right and wrong depend on the mere good pleasure of our Maker, who created us in a certain way. But if he had pleased, he might have made us to be pleased or displeased by quite different actions, even actions contrary to those that now please and displease us. Thus, right and wrong would be only matters of taste, only a certain effect in us, and nothing in actions themselves.

For his part, Price is convinced that morality is equally unchangeable with all truth and that right and wrong are real characteristics of actions and not mere sensations derived from the particular way in which our minds are framed. To show the immutability of right and wrong, Price argues that these ideas are derived not from a special sense but from the understanding. As Price sees it, the only debatable issue in morals is not what actions are right and wrong but what is the faculty by which we discern right and wrong.

Price prefaces his argument for regarding the understanding as our moral faculty with the preliminary claim that the understanding is a source of new ideas. He objects to interpreting John Locke as saying that sensation and reflection are the sources of all our ideas. Price argues that Locke may have meant only that all our ideas are ultimately grounded on ideas derived from sensation and reflection. Thus, Price makes room for certain new ideas that may arise as the understanding compares the objects of thought and judges them. Some of these new ideas are solidity, inertia, substance, accident, duration, space, cause or power, entity, possibility, and actual existence.

Price locates these new ideas in a revised classification of simple ideas. He divides simple ideas into those implying nothing real outside the mind and those that denote real and independent existence distinct from sensation. The first class of simple ideas consists, on the one hand, of tastes, smells, and colors and, on the other, of such notions as order, happiness, and beauty. The second class of simple ideas has three subclasses: the real properties of external objects, such as figure, extension, and motion; the actions and passions of the mind, such as volition, memory, and so on; and those new ideas noted above which arise as the understanding considers the ideas it has been supplied with. It is important to note that Price does not regard the second class of simple ideas as constructions of the mind. The real properties of external objects are in the objects, and such new ideas as cause, duration, and space are of properties in a real world.

Armed with his reclassification of simple ideas, Price is now prepared to locate our ideas of moral right and wrong in the scheme and thus establish that they are perceptions of the understanding. Price first considers the question of whether moral right and wrong are simple ideas. He declares that they must be, for we cannot give definitions of them that are more than synonymous expressions. It is Price’s recognition of this point which has led contemporary students to declare him one of the first to recognize the naturalistic fallacy, although he does not use that term. Having established that our ideas of right and wrong are simple ideas, Price then locates them in his scheme as two of those new ideas which arise in the understanding.

Hutcheson had simply assumed that if right and wrong are immediately perceived, they must be perceptions of an implanted sense. But the question of how we perceive these ideas may be settled by simply considering the nature of our own perceptions.

Let anyone compare the ideas arising from our *powers of sensation*, with those arising from our *intuition of the natures of things*, and enquire which of them his ideas of right and wrong most resemble.... It is scarcely conceivable that anyone can impartially attend to the nature of his perceptions, and determine that when he thinks gratitude or beneficence to be *right*, he perceives nothing *true* of them, and *understands* nothing, but only receives an impression from a sense.

Price notes that some impressions of pleasure or pain, satisfaction or disgust, generally attend our perceptions of moral right and wrong; the proponents of a moral sense may have confused these impressions with our actual perceptions of right and wrong.

But there is an assumption in Price's own system on which much depends and for which he offers insufficient argument. He tells us that "all actions undoubtedly have a *nature*. That is, *some character* certainly belongs to them, and somewhat there is to be *truly* affirmed of them." It is the task of the understanding to perceive these truths. Price regards actions in this way because it enables him to say that their rightness or wrongness is in them, not in the mind of the person judging the actions, but apart from noting the advantage to his own moral philosophy, Price offers no justification for the claim that actions have natures. It is unfortunate that he does not, for he rests his contention that morality is eternal and immutable on this claim.

When Price turns to our ideas of the beauty and deformity of actions, the second kind of perception of actions which he promised to account for, he finds that these perceptions are feelings of delight or detestation which may accompany our perceptions of the rightness or wrongness of actions. These feelings of delight and detestation are the effects on us of the actions we consider, and it is very likely that they arise from an arbitrary structure of our minds, which may be called a sense. Price allows that there is a distinction between noting that an action is right and approving it. We are made, however, in such a way that we cannot perceive an action to be right without approving it, for in humans it is necessary that the rational principle, or the intellectual discernment of right and wrong, should be aided by instinctive determinations. When these feelings of the heart support the perceptions of the understanding, we are provided with the motivation for moral behavior. Here Price agrees with Hutcheson, pointing out that he has never disputed that we owe much to an implanted sense and its determina-

tions. He means to resist only the claim that we owe our knowledge of right and wrong to such a sense.

Our ideas of the good and ill desert, the third sort of perception concerning actions which Price notes, carry the mind to the agent. He finds that we cannot but love a virtuous agent and desire his happiness above that of others. Quite apart from any advantage which we may gain from someone else's virtuous behavior, we have an immediate approbation of making the virtuous happy and of discouraging the vicious.

Price distinguishes between abstract and practical virtue. Abstract virtue denotes "what an action is independently of the sense of an agent; or what, in itself and absolutely, it is right such an agent, in such circumstances, should do." But Price recognizes that the actual practice of virtue depends on the opinion of the agent concerning his actions. Thus, an agent may be mistaken about his circumstances but sincere about what he believes he ought to do. In this respect practical virtue may diverge from abstract virtue but be no less obligatory insofar as the agent acts from a consciousness of rectitude. The ideal state of affairs is a correspondence of practical virtue with abstract virtue. Its achievement depends on the liberty and intelligence of the agent. These constitute the agent's capacity for virtue, and intention gives virtue actual being in a character. Price takes a short way with the question, "Why be moral?" "The knowledge of what is right, without any approbation of it, or concern to practise it, is not conceivable or possible. And this knowledge will certainly be attended with *correspondent, actual practice*, whenever there is nothing to oppose it." Why a person chooses to do what he knows he should do is a question "which need not and should not be answered."

Benevolence is not the sole virtue. We also have duties to God and to ourselves, and there is room for many other sorts of good behavior, such as veracity, sincerity, and gratitude. As a measure of virtue Price offers the rule that "the virtue of an agent is always less in proportion to the degree in which natural temper and propensities fall in with his actions, instinctive principles operate, and rational reflexion on what is right to be done, is wanting."

Price discusses at length the relation of morality to the divine nature. Just as moral right and wrong are independent of man's mind, they are also absolutes for God. Were this not so, there would be no sense in which God's will could be good.

## FREEDOM OF THE WILL

Price and Priestley published a set of letters as *A Free Discussion of the Doctrines of Materialism and Philosophical Necessity* (London, 1778). The correspondence had its origin in Price's criticism of Priestley's *Disquisitions Relating to Matter and Spirit*. The letters cover the nature of matter, the human mind, the mortality of the soul, the essence of the deity, and the doctrine of necessity. The last topic is the one that is treated in the most interesting way. Priestley contended that there can be no human liberty because "liberty" must mean someone's willing without a motive, which he regards as impossible. Price enlarges on the account of liberty that he offered in *A Review of the Principal Questions in Morals*. He argues that human agents are not physical objects but unique entities capable of self-determination. Consider the difference between a man who is dragged by a superior force and a man who follows a guide for a reward. Both of these examples may be certainties, but having different foundations, they are of totally different natures. "In both cases the man might in common speech be said to *follow*; but his following in the one case, however certain in event, would be *his own* agency: In the other case, it would be the agency of another. ... In the one case, superior power moves him: In the other he moves himself."

**See also** Burke, Edmund; Clarke, Samuel; Hume, David; Hutcheson, Francis; Liberty; Locke, John; Moral Principles: Their Justification; Moral Sense; Priestley, Joseph; Properties; Responsibility, Moral and Legal.

**Bibliography**

Price's *Works* were published in 10 volumes (London, 1816), with a memoir of his life by W. Morgan. *A Review of the Principal Questions in Morals* has been published with a critical introduction by D. D. Raphael (Oxford, 1948). This is a reprint of the third edition (1787) with an appendix and "A Dissertation on the Being and Attributes of the Deity."

For biography, see Carl B. Cone, *Torchbearer of Freedom, the Influence of Richard Price on Eighteenth Century Thought* (Lexington: University of Kentucky Press, 1952). Other works on Price include Joseph Priestley, *A Discourse on the Occasion of the Death of Dr. Price* (London, 1791); Leslie Stephen, *History of English Thought in the Eighteenth Century* (London, 1876; 2nd ed., London, 1902); and Roland Thomas, *Richard Price* (London: Oxford University Press, 1924).

*Elmer Sprague (1967)*

## PRIESTLEY, JOSEPH

(1733–1804)

Joseph Priestley, the English scientist, nonconformist minister, educator, and philosopher, was born at Birstall, Yorkshire, the son of a cloth dresser. His mother died in 1740, and in 1742 Priestley was adopted by a childless well-to-do aunt, Mrs. Keighley, a convinced but unbigoated Calvinist. A sensitive child, Priestley suffered greatly because he could not convince himself that he had experienced the "new birth" essential, on the Calvinist scheme, for his salvation. As a result of these childhood miseries Priestley was left, he tells us, with "a peculiar sense of the value of rational principles of religion" as opposed to the "ignorance and darkness" of Calvinism.

Until the age of sixteen Priestley was educated at a conventional grammar school. For the next three years, his health being too poor for regular studies, he in large part educated himself, reading his way into mathematics, physics, and philosophy and undertaking the study of European and Middle Eastern languages. In 1752 his health improved and he entered Daventry Academy, a university-type institution set up by nonconformists because Oxford and Cambridge would not admit nonconformists to a degree.

At Daventry the emphasis was on free discussion, and the curriculum was considerably broader than at Oxford or Cambridge. Priestley was introduced to David Hartley's *Observations on Man* (1749) and was at once—and permanently—converted to Hartley's general outlook. The simplicity and generality of Hartley's associationist psychology appealed to Priestley's maturing scientific instincts; it provided a theoretical foundation for his belief in perfectibility through education; and it offered a psychological alternative to the doctrine of free will, which Priestley's reading of Anthony Collins's *Philosophical Inquiry concerning Human Liberty and Necessity* (1714) had already caused him to reject.

In 1755 Priestley entered the ministry, taking over a decaying congregation at Needham Market, Suffolk. Stammering and unorthodox, he was not a success as a minister. He moved in 1758 to a more sympathetic but equally impoverished congregation at Nantwich in Cheshire. In an attempt to increase his income he set up a school where, perhaps the first to do so, he taught experimental science with the help of an "electrical machine" and an air pump.

Appointed in 1761 as "tutor of the languages" at Warrington Academy in Lancashire, Priestley taught oratory, literary criticism, grammar, history, and law, as well



as languages. Characteristically, on all these latter topics Priestley developed ideas that he sooner or later published. *The Rudiments of English Grammar* (1761), many times reprinted, is typical of his innovating boldness, insofar as he tried to simplify English syntax by removing from it the complications introduced by classically trained grammarians. His *A Chart of Biography* (1765) and *A New Chart of History* (1769) were even more enthusiastically received; they won for him not only his sole academic distinction, the doctorate of laws of the University of Edinburgh, but also his fellowship of the Royal Society.

Priestley's days of relative isolation were now over. In 1762 he married an ironmaster's daughter, Mary Wilkinson, an intelligent woman with a sense of humor and considerable force of character—qualities she was to need in the years to come. His duties at Warrington left him free to visit London for a month each year, where he came into contact with an active group of scientists, philosophers, and political thinkers, including Benjamin Franklin and Richard Price. Franklin encouraged Priestley's project of writing a history of electrical experiments. The work that resulted, *The History and Present State of Electricity, with original Experiments* (1767), is a notable contribution to the history of science. Describing a number of important original experiments, it is also in some respects the most theoretically adventurous of Priestley's scientific works. It contains as well Priestley's reflections on the use of hypotheses in scientific procedures as a guide to experimentation.

## EDUCATION AND GOVERNMENT

Like many of his fellow dissenters, Priestley was greatly interested in educational reform. Education had, he thought, thus far concentrated unduly on the needs of the clergy. His *An Essay on a Course of Liberal Education for Civil and Active Life* (1765) is a plea for a curriculum that should be suitable for men of affairs, emphasizing history and public administration rather than the classical languages. Priestley did much to encourage the teaching of history in the nonconformist academies. A set of lectures that he delivered at Warrington (published in 1788 as *Lectures on History and General Policy*) provided not only the academies but also the new American colleges with a text suitable for their needs; it was, indeed, recommended even at Cambridge. It is a summary account of the main historical sources, with an emphasis on commerce, law, and administration, rather than a historical textbook of the ordinary kind.

Priestley's political theory was closely related to his interest in education and his experience as a member of a minority group. In an appendix to his *Essay on a Course of Liberal Education* he developed an argument against the introduction of a state system of education, which would inevitably, he thought, favor the *status quo* and produce a quite undesirable uniformity of conduct and opinion. Like John Stuart Mill after him, Priestley gloried in diversity; uniformity, he said, is "the characteristic of the brute creation."

These reflections were more fully worked out in *An Essay on the First Principles of Government* (1768), which bears the subtitle *On the Nature of Political, Civil and Religious Liberty*. For Priestley, the preservation of civil liberty was the crucial political issue. Deciding who should participate in government—who, that is, should possess political, as distinct from civil, liberty—was, he thought, a practical matter, to be settled by considering what groups in the community are most likely, if they possess political power, to act for the greatest happiness of the greatest number. Such groups remain entitled to power only as long as they continue so to act. Legislation, on Priestley's view, should be kept to the minimum. What that minimum is cannot be determined a priori but only as a result of political experiment. But we can see at once, Priestley thought, that legislation that restricts civil and religious liberty is bound to be against the interests of the community. Unlike most nonconformist upholders of toleration and unlike his master John Locke, Priestley was uncompromising on this point; he upheld unbounded liberty of expression even to atheists and Roman Catholics.

In Priestley's eyes, the noblest of occupations was that of the clergyman, not the lecturer, and in 1767 he accepted a call to Mill Hill, Leeds, a congregation to whom his religious views were exceptionally congenial. The years Priestley spent at Mill Hill were extremely important in his development; his salary, although small, sufficed for his needs, and his duties left him considerable leisure.

## UNITARIANISM

Priestley had long before abandoned both the doctrine of the atonement, on which he wrote critically in *The Scripture Doctrine of Remission* (1761), and orthodox Trinitarianism. Now he took what was to be the final step in his transition from Calvinism to Unitarianism. Christ, he argued, although the Messiah, was a man, and not even a perfect man. Priestley's subsequent theological writings were in large part an attempt to prove—most maturely in

his *History of Early Opinions concerning Jesus Christ* (1786)—that Unitarianism was the doctrine of the early church. He defended his unorthodoxies both against clerical attack, as in his *Letters to Dr. Horsley* (1783–1786), and, as in his *Letters to a Philosophical Unbeliever* (Pt. I, 1780; Pt. II, 1787), against those who, like Edward Gibbon, could not understand why Priestley did not make a complete break with Christianity. Priestley valued his theological writings above all his other work. A firm belief in Providence is everywhere evidenced in his writings. Few men have committed themselves so often and so absolutely to the doctrine that “all is for the best in the best of all possible worlds,” although he also believed that the future world could—and therefore would—be better.

### SCIENTIFIC ACHIEVEMENT

It was as a scientist that Priestley won his international reputation. He published in 1772 what was intended to be the second section of a general history of science, *The History of the Present State of the Discoveries relating to Vision, Light and Colours*; but this work, invaluable though it still is to historians of science, did not arouse a great deal of interest. Priestley therefore abandoned his large-scale historical project and concentrated instead on chemistry. His first chemical publication, in 1772, was of an unusually practical character: It described a method of producing “mephitic julep,” or soda water. But it was the paper “On Different Kinds of Air,” which he read in that same year to the Royal Society,” that at once established his reputation as a chemist. In 1774 he prepared the first edition of *Experiments and Observations on Different Kinds of Air*; this he republished in a series of editions, with important changes in contents, in method of organization, and even in title, until 1790.

By the end of that period Priestley, following up the work of Joseph Black and Henry Cavendish, had considerably enlarged our knowledge of the chemical properties of gases. He differentiated between nine gases, of which only three had previously been known to science, and described a method of collecting them. Of particular importance was his preparation of “dephlogisticated air” (oxygen), which he produced on August 1, 1774, by heating red mercuric oxide. It then became clear that air was not an element. Priestley went on to examine the properties of oxygen; in a series of chemicobiological experiments he brought out its importance for animal life.

As a resourceful experimenter, using simple and economical methods, Priestley has had few equals. But it was left to others, to Cavendish and Antoine Lavoisier, to appreciate the theoretical significance of his work. Priest-

ley had isolated oxygen and had observed its importance in combustion; he had passed a spark through a mixture of hydrogen and oxygen and had noticed that dew was formed. Yet his last scientific work (1800) bore the title *The Doctrine of Phlogiston established and that of the Composition of Water refuted*. Although he had himself carried out important quantitative experiments, he did not appreciate the significance of the quantitative considerations by which Lavoisier overthrew the phlogiston theory.

### PHILOSOPHY

Much of Priestley’s most important scientific work was carried out at Shelburne, where from 1772 until 1780 he acted as “librarian and literary companion” to the Earl of Shelburne. During these same years Priestley embarked upon his most substantial metaphysical works. He began in 1774 with *An Examination of Dr. Reid’s Inquiry into the Human Mind on the Principles of Commonsense, Dr. Beattie’s Essay on the Nature and Immutability of Truth, and Dr. Oswald’s Appeal to Commonsense on Behalf of Religion*, commonly referred to as *An Examination of the Scotch Philosophers*. This is a vigorous polemic, which sets out to demonstrate the superiority of Hartley’s psychology to the philosophy of the Scottish commonsense school, a philosophy that Priestley thought obviously reactionary insofar as it substituted for the simple Locke-Hartley theory of mind “such a number of independent, arbitrary, instinctive principles that the very enumeration of them is really tiresome.” All the so-called instinctive beliefs of common sense can, Priestley set out to show, be derived from the operations of associative principles working on the materials provided by sensation. He came to regret in later life the tone of this publication but never its doctrines.

**MATERIALISM.** Hoping to make Hartley’s views better known, Priestley published an abridged version of Hartley’s *Observations on Man* in 1775 as *Hartley’s Theory of the Human Mind on the Principle of the Association of Ideas*. In his preface, Priestley somewhat tentatively suggested that all the powers of the mind might derive from the structure of the brain. Even as a suggestion this created a considerable uproar, but Priestley was not to be intimidated by clerical clamor. Convinced that materialism was the natural metaphysical concomitant of Hartley’s associative psychology, he set out, therefore, in his *Disquisitions Relating to Matter and Spirit* (1777) to demonstrate that materialism was theologically, scientifically, and metaphysically superior to orthodox dualism.

On the theological side, materialism had commonly been objected to on the ground that it is incompatible with immortality. Man, Priestley replied, is not “naturally” immortal; he is immortal only because, as we know from revelation, God chooses to resurrect him; this resurrection is of the body and therefore also of the body’s mental powers. As for the commonplace metaphysical objections to materialism, these are based, according to Priestley, upon an untenable conception of matter as being by nature inert and therefore incapable of exerting mental activity. To such a concept of matter Priestley opposed the physical theories of his friend and fellow scientist John Michell and the Jesuit mathematician Roger Boscovich. Material objects, on their view, are centers of force; if this is the nature of matter, Priestley argued, there is no good reason for denying that mental operations are part of the activity of a material object. On the other hand, there are very good reasons for objecting to the traditional dualism, which is quite incapable of explaining how mind and body can enter into any sort of relationship.

**DETERMINISM.** Priestley had been a determinist long before he became a materialist, but not until 1777, in *The Doctrine of Philosophical Necessity Illustrated*, did he fully present his case against free will; indeed, even then he thought of himself as supplementing Thomas Hobbes, Collins, David Hume, and Hartley with illustrations rather than as working out an entirely independent position. The doctrine of free will, he argued, is theologically objectionable because it cannot be reconciled with the existence of an all-seeing Providence; from a metaphysical standpoint, it makes human actions quite unintelligible, and ethics has no need of it. As a basis for our everyday moral judgments, the distinction between acting voluntarily and acting under compulsion is certainly important, but this distinction does not, according to Priestley, rest upon a metaphysical conception of free will.

Priestley’s metaphysical unorthodoxies considerably disturbed his old friends, provoking a candid but good-tempered correspondence with Richard Price, published in 1778 as *A Free Discussion of the Doctrines of Materialism and Philosophical Necessity Illustrated*. Developing his views on the relation between moral judgments and determinism, Priestley admitted that the determinist will prefer to avoid describing people as blameworthy or praiseworthy. He will say of them, rather, that they have acted, or have not acted, from good principles—from principles, that is, that are conducive to the general happiness. But the determinist’s different method of describ-

ing moral conduct has, Priestley thought, no practical consequences, and if determinism is in some respects inconsistent with everyday usage, this is even more true of libertarianism.

## LATER YEARS

There was a real risk, however, that Priestley’s reputation for materialism might endanger the earl of Shelburne’s political ambitions. Perhaps for this reason Priestley and Shelburne parted amicably in 1780, when Priestley, refusing Shelburne’s offer of a post in Ireland, took up residence in Birmingham. There he had a circle of congenial friends who were prepared to offer him financial as well as intellectual support. He became a member of the Lunar Society, with which were associated men of the caliber of Erasmus Darwin and James Watt, and he enjoyed the friendship and help of the scientifically minded potter Josiah Wedgwood, who supplied him with apparatus specifically designed for his chemical experiments. Much of Priestley’s scientific work in this period, under Alessandro Volta’s influence, conjoined his two main scientific interests: electricity and gases. He examined the effect of passing electrical sparks through a variety of gases and studied their thermal conductivity.

He was by no means unsympathetic to the laissez-faire sociopolitical attitude of Birmingham industrialists. In *Some Considerations on the State of the Poor in General* (1787) he strongly criticized the poor laws and elsewhere opposed apprenticeship laws and laws for regulating interest rates. On his view, any sort of social welfare legislation “debased the very nature of man” by treating him as someone who had to be provided for. Although Priestley warmly supported schemes for cooperative insurance against hardship, he was opposed to any legislation that might diminish independence or increase the power of the state over individuals.

**POLITICAL RADICALISM.** In general terms, Priestley’s life at Birmingham was a continuation and development of his earlier activities; theological controversy continued to be his main interest. But one event transformed his life and modified his political attitudes: the French Revolution. Reacting to that revolution, the British government became steadily more intolerant and conservative, and Priestley came to think that extensive political innovations were a necessary condition for the preservation of civil liberty. He moved toward political radicalism of the nineteenth-century kind in his *Letters to Edmund Burke occasioned by his Reflections on the Revolution in France* (1791) and in the anonymously published *A Political Dia-*

*logue on the General Principles of Government* (1791). He had formerly been accustomed to describe himself as “a Unitarian in religion but a Trinitarian in politics” because he had accepted the view that liberty rested on the balance between king, Commons, and House of Lords. He now came to feel that there should be but one source of political power, the will of the people as it would be represented in a reformed House of Commons.

On July 14, 1791, the Friends of the Revolution organized a dinner at Birmingham (Priestley was not present) in order to commemorate the fall of the Bastille. This was the last straw. With the encouragement, it would seem, of the authorities, an angry mob attacked the non-conformist chapels, then turned their attention to Priestley’s house, destroying his books and furniture. Priestley was persuaded by his friends to leave Birmingham for London where he was, however, shunned by his scientific colleagues.

**LIFE IN AMERICA.** For some years, Priestley had been contemplating migration to the United States, where his three sons had already gone. In 1794 he left for New York and finally settled in Northumberland, Pennsylvania. There, still supported by his old friends, he continued to experiment and to write, mainly on theological questions.

He was disappointed, however, by the orthodoxy of the American clergy and alarmed by the growth of intolerance in the United States. Although he took no part in politics, he wrote an uncompromising exposition of his political and religious views in *Letters to the Inhabitants of Northumberland* (1799). There was talk of his being deported under the Aliens Act, but John Adams would not permit the application of the act to “poor Priestley.” With the election of Thomas Jefferson to the presidency, Priestley was not only secure but also at last on good terms with authority. Jefferson consulted him on educational questions, and Priestley’s *Socrates and Jesus Compared* (1803) precipitated Jefferson’s “Syllabus” of his religious beliefs. Another of Priestley’s works, *The Doctrines of Heathen Religion Compared with those of Revelation* (1804), awoke in Adams an enthusiasm for comparative religion. Priestley’s last years, from 1801 until his death, were marred by ill health and bereavements, but his diversified intellectual interests remained with him until the end.

**See also** Boscovich, Roger Joseph; Collins, Anthony; Darwin, Erasmus; Determinism, A Historical Survey; Determinism and Freedom; Franklin, Benjamin; Hartley, David; Hobbes, Thomas; Hume, David; Jefferson,

Thomas; Libertarianism; Locke, John; Materialism; Mill, John Stuart; Price, Richard.

### **Bibliography**

Priestley’s scientific writings have never been collected. The standard edition of his philosophical, theological, and miscellaneous writings is *The Theological and Miscellaneous Works of Joseph Priestley*, edited by John Towill Rutt, 25 vols. in 26 (London: G. Smallfield, 1817–1832). John Arthur Passmore, *Joseph Priestley* (New York, 1965) includes a short list of Priestley’s writings, an introductory essay, and selections from his major works.

See also Thomas Henry Huxley, “Joseph Priestley,” in *Science and Culture* (London, 1881); Thomas Edward Thorpe, *Joseph Priestley* (New York: Dutton, 1906); Edgar Fahs Smith, *Priestley in America, 1794–1804* (Philadelphia: P. Blakiston’s Son, 1920); Wallace Ruddell Aykroyd, *Three Philosophers: Lavoisier, Priestley and Cavendish* (London: Heinemann, 1935); Arthur Handley Lincoln, *Some Political and Social Ideas of English Dissent, 1763–1800* (Cambridge, U.K., 1938); Francis Edward Mineka, *The Dissidence of Dissent* (Chapel Hill: University of North Carolina Press, 1944); Stephen Edelston Toulmin, “Crucial Experiments: Priestley and Lavoisier,” in *Journal of the History of Ideas* 18 (2) (1957): 205–220; and the detailed bibliographies in the *Dictionary of National Biography* and the *Dictionary of American Biography*.

*John Passmore (1967)*

## PRIMARY AND SECONDARY QUALITIES

The distinction between “primary and secondary qualities,” first stated and thus named by Robert Boyle, received its classical formulation in John Locke’s *Essay*. There Locke states that apart from ordinary causal properties or “powers,” material objects possess five primary qualities—extension (size), figure (shape), motion or rest, number, and solidity (impenetrability)—and many secondary qualities, such as color, taste, smell, sound, and warmth or cold. This distinction was made in the context of representative realism; that is, it was presupposed that the qualities of objects are quite distinct from, and are in fact causes of, “ideas” (representations or *sensa*), which are the only immediate objects of sensory awareness. The basis of the distinction was twofold. First, perceived size, shape, motion, number, and solidity are ideas caused by and exactly resembling the corresponding primary qualities of objects; perceived color, taste, smell, sound, and so on are caused by, but do not resemble, the corresponding secondary qualities. Second, the primary qualities are inseparable from matter and are found in every part of it;

the secondary qualities are not true qualities of matter but are merely powers in the objects to produce sensory effects in us by means of the primary qualities in their minute parts. Thus, red as experienced (*idea* or *sensum*) is the effect of the secondary quality red, which is merely the power possessed by a special texture or surface structure of the object to reflect certain light frequencies and to absorb others.

This formulation is rather clumsy, and since George Berkeley the custom has been to apply the first part of the distinction to the qualities of the ideas or *sensa*. The primary or spatiotemporal qualities of these data may then be said to characterize the object as well, for instance, the *sensum* is square and so is the object; but the secondary qualities are said not to characterize the object at all except in a derivative way, for instance, the *sensa* may be red and fragrant, but the object itself is intrinsically neither colored nor scented; it is red and fragrant only in the secondary sense that it causes the appropriate data of color and smell in the percipient. The doctrine is thus essentially the same as Locke's, but the language is slightly different. This second formulation will be used here.

Though Boyle and Locke invented and popularized the distinction and the terminology of primary and secondary qualities, the distinction dates back in principle to Democritus, who said that sweet and bitter, warm and cold, and color exist only by convention (*νόμος*), and in truth there exist only the atoms and the void (Fr. 9, Diels and Kranz). The distinction was revived by Galileo Galilei and accepted by René Descartes, Isaac Newton, and others.

## ARGUMENTS FOR THE DISTINCTION

**RELATIVITY AND MEASUREMENT.** The relativity argument is the most important one: Secondary qualities are affected by the condition of our sense organs and nervous system, by our distance from the object or its motion relative to us, by the lighting or by such intervening media as fog. Since secondary qualities thus vary according to, and depend for their nature on, factors quite external to the physical object, they cannot be intrinsic properties of it. This point was elaborated by Locke in various examples, two of which follow: (a) If one takes three bowls of water, one judged hot, one judged cold, and one judged medium, and places one hand in the hot water and the other hand in the cold, and then transfers both hands to the middle bowl, the water in that bowl will feel hot to the hand that has been in the cold water and cold to the hand that has been in the hot water. But since it cannot be both hot and cold, hot and cold are therefore not intrinsic

properties of the water. (b) Marble is not colored in the dark; its color appears only in the light. But presence or absence of light cannot alter its real properties, so that the perceived color cannot be included among them.

If we grant the position of representative realism that hot, cold, and color, as experienced, are qualities of ideas or representations, then it is plausible to suppose on these grounds that they do not also characterize objects or resemble properties of objects. (Locke does not always make it clear that representative realism is to be presupposed). But this claim is apparently open to the insuperable objection, stated by Berkeley, that the primary qualities also vary: The object's apparent shape or size varies just as much as its color or sound. This would mean that shape and size as perceived do not characterize objects or resemble the actual properties of the object, thus subverting the whole basis of the distinction. That Locke did not see this may have been partly because he felt that he had to argue against the commonsense assumption that all sensible qualities characterize objects, and partly because the belief that primary qualities characterize all matter was apparently guaranteed by the physics of his day.

Although this objection is valid against Locke's position, it does not destroy the distinction between the primary and secondary qualities, which it is natural to recast and support by a revised relativity argument. This new point is that, in contrast with the secondary qualities, the main primary qualities—shape, size, and motion—can all be measured (solidity cannot, but it is dubious anyhow, in that most physical objects, even atoms, are far from solid or impenetrable; number, whether there is one object or two, seems scarcely a quality at all; strictly also in the case of shape, what is measured are various dimensions—diameters, angles, and so on—of the object, and supporters of the distinction must maintain that these are the *differentiae* of the shape). A plate may look elliptical, but by measuring its diameters and seeing that they are equal, we can establish that it is round; one man may look taller than another, but their relative heights may be settled by measurement, as can the speed of objects relative to the earth. The measured size and shape of a plate may thus be held to characterize it, and the sensible size and shape may agree with and resemble them, so that one can say that size and shape (and motion) are primary. Nevertheless, only in favorable circumstances does a given primary sensible quality also characterize the object (for instance, both object and *sensum* are round); otherwise, there is only a projective relationship, as between elliptical *sensum* and round object.

Measurement is objective and does not vary significantly because it is an operation that depends on the coordination of a number of separate perceptions and that may be performed by a number of different persons. Consequently, variations due to the measurer on any particular occasion are compensated for and do not affect the final result, and the various actions confirm that one is not simply establishing the qualities of representations. Measurement also leads to conclusions regarding the dimensions and positions of objects in physical space that can be verified by further activities or operations, such as fitting the objects together, moving one's hand between them, rolling an object to confirm that it is round, and so on. By contrast, the variation found in the sensory qualities seems to be caused by their being simply the content of one single act of perception limited to one person at one time.

If all this is so, the list of primary qualities must be somewhat amended. Shape, size, and motion remain, but one should substitute mass for solidity. Temperature is more difficult: Since it can be measured, it seems at first primary. But what is measured is the property of causing expansion in fluid or metal; this property in no way resembles felt warmth, and in physical theory it is a form of energy. Hence, temperature should not be regarded as a separate primary quality. Material objects do, of course, possess many other properties—causal and dispositional ones, for example—as Locke realized by his doctrine of “powers,” but part of the distinction is that only the primary ones are intrinsic (that is, possessed without reference to other objects) and that all such powers are ultimately due to patterns of primary qualities. Even so, the distinction would have difficulty in coping with some intrinsic “scientific” properties, such as energy or electric charge.

Apart from this, various objections have been made to the distinction in terms of measurement. First, measured motion and size must be stated in terms of some standard, such as a yard or meter; hence, they are purely relational and are not intrinsic properties of the object. But one can reply that it is only the description or labeling of the measurement that is thus relational; the motion or extension labeled, which is actually measured, seems intrinsic to the object.

Second, since colors and sounds may be measured, are they not also primary? But this objection seems based on a misunderstanding of the processes of measurement, for one way of “measuring” color might be to compare a given shade with a standard on a shade card; but that would be like comparing the sensible size of two objects, not meas-

uring them. Proper measurement goes beyond this kind of sensory experience, and even if one gives the shade a number, one cannot calculate with the results as one can with the dimensions of objects. Normally, however, measurement of colors or sounds is either the measurement of the amplitudes or lengths of light waves or sound waves, or a mixture of wave measurement and the comparison of experiences. If one brings up a decibel meter and says that the sound to be measured is 80 decibels, it is the amplitude of the sound waves that is ultimately responsible for the movement of the pointer to 80. It should be noted, however, that the logarithmic scale is used because of a characteristic of human ears—that experienced loudness is related logarithmically to wave amplitude.

Third, measurement is a perceptual process—at least it relies on and largely uses perception—so it may be only producing various correlations of *sensa* and never getting through to the supposed properties of material objects at all. This objection is made from the point of view of phenomenalism, however, while the whole primary-secondary quality distinction presupposes representative realism. Supporters of the latter would say that the best explanation of the correlation is that the sense experiences arise in the measurement of actual physical objects.

**ARGUMENTS FROM SCIENCE.** Science can adequately explain and describe the nature of the physical world solely in terms of primary qualities; hence, while primary qualities must characterize objects, there is no need to suppose that secondary qualities must also. The latter would be otiose, and on the principle of economy, or Ockham's razor (that entities should not be multiplied more than is necessary), it would be unscientific to suppose that they exist as intrinsic properties of objects. The objection to this argument is partly that the science of one's day is not final (thus, Locke was persuaded by seventeenth-century science to include solidity in the list of primary qualities), and mainly that scientific theory and description are not the whole truth—they describe only one aspect of the world, being limited by their quantitative approach and their instruments. Secondary qualities may thus be real properties of matter with biological or aesthetic functions; Ockham's razor oversimplifies the facts pertaining to living things.

Investigation of the causal processes on which perception depends shows that the only variables capable of transmitting information about the properties of external objects are spatiotemporal ones, which are associated with primary qualities. Thus, light waves (energy distributed in space and time) pass from the object to the per-

ipient, but nothing resembling experienced color and sound is transmitted. But the main force of this argument, since it applies to all the senses, is neurological. The nerves from the different sense organs to the brain are all similar, and therefore the only variables are the frequencies of the impulses (which convey the intensity of the stimulus), their different neural pathways, and their different destinations in the brain. Indeed, it seems to be the different destinations that primarily govern the type and quality of the sensation. And although one can conceive of primary qualities being transmitted by spatiotemporal variables, it is difficult to conceive of color, warmth, taste, or smell being so transmitted. (It may be objected that radio and television can transmit color and sound by converting them into electrical impulses for transmission and then reconverting them. But, strictly speaking, what is converted is not color or sound but light waves or sound waves; moreover, the radio or television station must use microphones and cameras to effect the conversion, and there is no evidence of such conversion devices at the objects we see or hear.)

### BERKELEY'S CRITICISMS

Berkeley's formidable criticisms of the distinction between primary and secondary qualities have convinced many people. We have mentioned his objection concerning relativity, which, though valid against Locke, can be avoided by restating the distinction on the basis of measurement. He also has nothing to say on the scientific considerations, which were not explicit in Locke. But he did have some further well-known criticisms. First, he stated, "An idea can be like nothing but an idea." In other words, our *sensa*, being private, mental, and directly perceivable, cannot resemble properties of material objects that are public, physical, and not objects of direct awareness. But resemblance is claimed only for primary qualities; and though *sensa* cannot be extended in physical space, it seems reasonable to claim a structural resemblance, a similarity in form, between the spatial relations that they sensibly possess and those attributed to objects by measurement; thus, it can be confirmed by measurement that various relations between the sides of a square *sensum* hold in the object. A similar resemblance seems plausible in the case of motion. There are, however, some underlying difficulties here. In the older representative realism, *sensa* were mental; and since the mind was held to be unextended, they could hardly have spatial relations. But newer versions would allow some sensible or subjective space different from physical space; certainly *sensa* seem spatial, and there seems to be no reason why what is

directly perceivable and what is not should be unable to have a similar form or character.

Second, matter consisting only of primary qualities—for instance, possessing extension but no color, taste, sound or smell—is inconceivable. This objection is beside the point: Admittedly one cannot conceive, in the sense of "imagine" or "picture to oneself" (Berkeley's sense of the word), any such thing, for what we can imagine is limited by past experience and perception. But the range of possible existents need not be confined to this, and there is much in science, particularly in modern physics, that cannot be imagined or pictured.

**See also** Berkeley, George; Boyle, Robert; Colors; Descartes, René; Galileo Galilei; Leucippus and Democritus; Locke, John; Newton, Isaac; Pain; Perception; Realism; *Sensa*; Sound.

### Bibliography

#### CLASSICAL WORKS

Anticipations of the distinction are to be found in Democritus, Fragments 9 and 125, in H. Diels and W. Kranz, *Fragmente der Vorsokratiker*; Galileo, *Il saggiaiore* (The Assayer; 1623), in his *Opere*, edited by F. Flora (Milan, 1952), pp. 311–314; René Descartes, *Principles of Philosophy*, Part IV, Nos. 188–203, and *The World or Essay on Light*, Chs. 1–2. Robert Boyle adumbrated the distinction in his *The Origin of Forms and Qualities* (London, 1666), especially the section titled "An Excursion about the Relative Nature of Physical Qualities"; John Locke stated it in his *Essay concerning Human Understanding* (London, 1690), Book II, Ch. 8; Newton's view is best stated in his *Opticks*, Book I, Part II (paperback, New York, 1952, p. 124), but that was not published until 1704. Berkeley's criticisms are in his *Principles of Human Knowledge*, Secs. 9–20. Edwin A. Burtt, *The Metaphysical Foundations of Modern Physical Science* (London: K. Paul, Trench, Trubner, 1925), contains useful discussion of the early versions.

#### MODERN WORKS

No book is devoted to the distinction, and substantial treatments are rare, but helpful contributions may be found in the following:

A more recent scientific comment on the distinction is Russell Brain, *Mind, Perception and Science* (Oxford: Blackwell, 1951). L. Susan Stebbing, *Philosophy and the Physicists* (London: Methuen, 1937), attacks scientific versions of it, especially those of Newton and of Arthur S. Eddington (see his *The Nature of the Physical World*, Cambridge, U.K.: Cambridge University Press, 1928).

Among the general works available are Charlie Dunbar Broad, *Scientific Thought* (London: Kegan Paul, 1923) and *The Mind and Its Place in Nature* (London: Kegan Paul, 1925); William Pepperell Montague, *The Ways of Knowing* (London: Allen and Unwin, 1925); Roderick Chisholm, *Perceiving: A Philosophical Study* (Ithaca, NY: Cornell University Press, 1957); and more fully, R. J. Hirst, *The*

*Problems of Perception* (London: Allen and Unwin, 1959). D. M. Armstrong, *Perception and the Physical World* (New York: Humanities Press, 1961), Chs. 14 and 15, discusses the subject from a direct realist position, and Gilbert Ryle's criticisms of the distinction in his *Dilemmas* (Cambridge, U.K.: Cambridge University Press, 1954) are well known.

Among articles on the distinction are Reginald Jackson, "Locke's Distinction between Primary and Secondary Qualities," in *Mind* 38 (149) (1929): 56–76; J. J. C. Smart, "Colours," in *Philosophy* 36 (137) (1961): 128–142 (reprinted with additions in his *Philosophy and Scientific Realism*, New York: Humanities Press, 1963); Colin Strang, "The Perception of Heat," in *PAS* 61 (1960–1961): 239–252; and for a more general treatment, W. C. Kneale, "Sensation and the Physical World," in *Philosophical Quarterly* 1 (2) (1951): 109–126.

#### OTHER RECOMMENDED TITLES

- Averill, Edward Wilson. "The Primary Secondary Quality Distinction." *Philosophical Review* 91 (1982): 343–362.
- Ayers, Michael. *Locke: Epistemology and Ontology*. London: Routledge, 1991.
- Bennett, Jonathan. *Locke, Berkeley, Hume: Central Themes*. Oxford: Oxford University Press, 1973.
- Cummins, Robert. "Two Troublesome Claims about Qualities in Locke's *Essay*." *Philosophical Review* 84 (1975): 401–418.
- Dicker, Georges. *Perceptual Knowledge*. Dordrecht: Reidel, 1980.
- Fales, Evan. *A Defense of the Given*. Lanham, MD: Rowman & Littlefield, 1996.
- Fumerton, Richard. *Metaphysical and Epistemological Problems of Perception*. Lincoln: University of Nebraska Press, 1985.
- Jackson, Frank. *Perception: A Representative Theory*. Cambridge, U.K.: Cambridge University Press, 1977.
- Levin, Janet. "Physicalism and the Subjectivity of Secondary Qualities." *Australasian Journal of Philosophy* 65 (1987): 400–411.
- Mackie, John. *Problems from Locke*. Oxford: Clarendon Press, 1976.
- McGinn, Colin. *The Subjective View*. Oxford: Clarendon Press, 1983.
- Sellars, Wilfrid. *The Metaphysics of Epistemology*. Atascadero, CA: Ridgeview Press, 1989.
- Smith, A. D. "Of Primary and Secondary Qualities." *Philosophical Review* 99 (2) (1990): 221–254.
- Wright, Crispin. "Moral Values, Projection, and Secondary Qualities." *PAS*, Supp. 62 (1988): 1–26.

**R. J. Hirst (1967)**

*Bibliography updated by Benjamin Fiedor (2005)*

## PRINGLE-PATTISON, ANDREW SETH (1856–1931)

Andrew Seth Pringle-Pattison, the Scottish personal idealist, was born Andrew Seth, in Edinburgh. (He adopted the surname Pringle-Pattison at the age of forty-two as a

condition of inheriting a family estate in Scotland.) He studied philosophy at Edinburgh University under Campbell Fraser. Two years of study in Germany convinced him that it was the worst place for the study of German idealism but resulted in his completing, at twenty-four, his Hibbert essay, *The Development from Kant to Hegel*. From 1880 to 1883 he served as Fraser's assistant at Edinburgh and then took the foundation chair of philosophy in the University College of South Wales at Cardiff. He left Cardiff in 1889 for the chair of logic and metaphysics at the University of St. Andrews. This he relinquished in 1891, when he succeeded Fraser at Edinburgh. In 1919 he resigned, after thirty-nine influential years as a university teacher.

Philosophy for Pringle-Pattison was a serious enterprise of the human spirit, which he did not distinguish strictly from a statement of his own findings in religion and morality. His writing is clear and eloquent but not very original. He sought to advance his subject through critical interpretation of the great philosophers, especially Immanuel Kant and G. W. F. Hegel. He was skeptical about the value of philosophical systems, holding that we cannot know the universe as we can know its individual parts; only God can do this. Rather, "the ultimate harmony may justifiably be spoken of as an object of faith—something which I am constrained to believe, even though I do not fully see it."

Pringle-Pattison was a Scottish Hegelian with a difference. Rebellious against the absolutism of Hegel and of such Hegelians as Francis Herbert Bradley and Bernard Bosanquet, for whom the individual is merged in the universal, he insisted on the uniqueness of the individual person. It is only as knower that the self is a unifying principle. As a real being it is separate and distinct, impervious to other selves, even to God. "I have a centre of my own—a will of my own—a centre which I maintain even in my dealings with God Himself." We feel this to be so; it neither needs to nor can be established by argument. But God too is a Person; we cannot deny him self-consciousness, because this is the highest source of worth in ourselves. Hegel and the Hegelians were at fault here also.

Philosophy, Pringle-Pattison held, cannot do justice to "the individual within the individual—those memories, thoughts, and feelings which make each of us a separate soul" (*Hegelianism and Personality*, p. 217). Religion and poetry go further and deeper than philosophy, and this, as he said, is why he drew so frequently on the poets.

Our knowledge of the Absolute starts from experience—our experience "of the concrete worlds of morality, of beauty, of love or of the passion of the intellectual life."



It is, however, a postulate of reason that the world is a cosmos, not a chaos, which we can gradually explore but never grasp in its entirety. Pringle-Pattison described his philosophy as “a larger idealism” that reconciles the dictates of morality and religion with the findings of science, purpose being the supreme category.

He was cautious in his claims about immortality. The nature of the soul is such that it is reasonable to entertain the hypothesis of its survival, and since human spirits must be “values for God” they were surely not made to be constantly destroyed and replaced by others. Yet if there is personal immortality, it is not the inherent possession of every human soul but must be won by the continuous effort needed to develop a coherent self. Morality does not depend on personal immortality, nor need immortality be the central article of philosophy or religion. In the apprehension of Truth, Beauty, and Goodness—eternal realities—man has already tasted eternal life and so should not be much concerned about personal survival.

**See also** Absolute, The; Bosanquet, Bernard; Bradley, Francis Herbert; Hegel, Georg Wilhelm Friedrich; Idealism; Kant, Immanuel.

## Bibliography

### WORKS BY PRINGLE-PATTISON

*Scottish Philosophy: A Comparison of the Scottish and German Answers to Hume*. Edinburgh: Blackwood, 1885.

*Hegelianism and Personality*. Edinburgh: Blackwood, 1887.

*Man's Place in the Cosmos*. Edinburgh: Blackwood, 1897.

*The Idea of God in the Light of Recent Philosophy*. Oxford: Clarendon Press, 1917.

*The Idea of Immortality*. Oxford: Clarendon Press, 1922.

*Studies in the Philosophy of Religion*. Oxford: Clarendon Press, 1930.

*The Balfour Lectures on Realism* (delivered 1891), with a memoir by editor G. F. Barbour. Edinburgh: Blackwood, 1933.

**A. K. Stout (1967)**

## PRIOR, ARTHUR NORMAN (1914–1969)

Arthur Norman Prior was born on December 4 at Masterton, near Wellington, New Zealand. He acknowledged an early philosophical debt to John Findlay. But his first academic post was at Canterbury University College, where he succeeded Karl Popper. He was the visiting John Locke Lecturer at Oxford in 1956, and in 1958 he was appointed a professor of philosophy at the University of

Manchester. After short periods as a visiting professor at the University of Chicago and at the University of California at Los Angeles, he moved in 1966 to a tutorial fellowship at Balliol College, Oxford, and Oxford University appointed him to a concurrent readership.

Prior's early intellectual interests were very much religious in character. He was influenced for several years by the theologian Arthur Miller, who combined a strict adherence to Presbyterian doctrine with an equally strong support for socialism and opposition to nationalism. But Prior's pacifism weakened, and he served from 1942 to 1945 in the New Zealand air force. And the central focus of his interests gradually shifted—helped by an occasional bout of atheism—from theology to ethics and logic. He exchanged ideas with a wide circle of friends and acquaintances, and his hospitality to students was legendary.

Prior's first book, *Logic and the Basis of Ethics* (1949) traced seventeenth-, eighteenth-, and nineteenth-century anticipations of G. E. Moore's criticism of the so-called naturalistic fallacy. But his main claim to fame lies in his pioneering work on the formal logic of temporal relationships. His most important investigations in this field were published in *Time and Modality* (1957), *Past, Present, and Future* (1967), and *Papers on Time and Tense* (1968). But he also wrote on several logical topics in this encyclopedia; he published a substantial survey of the current state of logical inquiry under the title of *Formal Logic* (1955; 2nd ed., 1962); and a posthumous volume of papers, *Objects of Thought* (1971), was edited by P. T. Geach and A. J. P. Kenny.

Prior almost always used the Polish style of notation in the discussion of logical proofs and principles and was a convinced, though largely unsuccessful, champion of its virtues. The major inadequacy in his tense logic, however, was a failure to discuss or accommodate aspectual differences—roughly, differences between the meanings expressed by verbs in a perfect tense and those expressed by verbs in an imperfect tense (see Galton, 1984). Other criticisms may be found in L. J. Cohen's (1958) review of *Time and Modality* and in his subsequent controversy with Prior (*Philosophy* 34 [1959]). In his *Formal Logic* Prior displayed an impressively wide acquaintance with logical systems outside the field of tense logic, and this book remains a useful text for anyone interested in comparisons between different axiomatizations of the propositional calculus, between different kinds of logical quantification, between different modal logics, or between different three-valued or institutionist logics. But the treatment of metalogical issues in the book is

occasionally rather selective: for example, in its discussion of completeness proofs for the predicate calculus as against its treatment of completeness proofs for the propositional calculus.

Outside the brilliant originality of his work on tense logic, perhaps Prior's most striking idea was expressed in "The Runabout Inference-Ticket" (1960), where he argued that, if the meaning of a logical connective consisted just in the logical uses to which it can be put (as many seemed to hold), then it would be easy to invent a connective with a meaning that would enable one to infer any conclusion from any premises.

*See also* Atheism; Modal Logic; Moore, George Edward; Nationalism; Pacifism; Popper, Karl Raimund; Socialism.

### **Bibliography**

#### WORKS BY PRIOR

- Logic and the Basis of Ethics*. Oxford: Clarendon Press, 1949.  
*Formal Logic*. Oxford: Clarendon Press, 1955. 2nd ed., 1962.  
*Time and Modality*. Oxford: Clarendon Press, 1957.  
 "The Runabout Inference-Ticket." *Analysis* 21 (1960): 38–39.  
*Past, Present, and Future*. Oxford: Clarendon Press, 1967.  
*Papers on Time and Tense*. Oxford: Clarendon Press, 1968.  
*Objects of Thought*, edited by P. T. Geach and A. J. P. Kenny. Oxford: Clarendon Press, 1971.

#### WORKS ON PRIOR

- Cohen, L. J. Review of *Time and Modality*. *Philosophical Quarterly* 8 (1958): 266–271.  
 Gallon, A. *The Logic of Aspect*. Oxford: Clarendon Press, 1984.

*L. J. Cohen (1996)*

## PRIORITY OF KNOWLEDGE, THE

*See Knowledge, The Priority of*

## PRISONER'S DILEMMA

*See Decision Theory*

## PRIVATE LANGUAGE PROBLEM

The private language problem is essentially the question of whether or not a language as a system of symbols that

are means of thinking is, of necessity, a language as a system of symbols that are means of communication. Defining "private language" as language (in the sense of means of thinking) which in principle the speaker alone can understand (so that it cannot serve as a means of communication), our question is roughly equivalent to: "Is a private language possible?" Many philosophers, following Ludwig Wittgenstein, have made the claim (here called the private language thesis, abbreviated PLT) that private languages are impossible. Armed with it, they have argued against solipsism, phenomenism, the analogical or empirical view of one's knowledge of other minds, and against mind-body dualism. Some of them have gone on to argue for certain versions of philosophical behaviorism as well as for the view that the meaning of a word consists of its use or employment in a social practice and not in its referring to something or its designating a kind of entity.

Thus, the PLT has been a central principle in the cluster of Wittgensteinian doctrines. It is not clear, however, that exactly the same thesis figures in all the arguments in question, since the idea of a private language varies in different contexts. There is, therefore, a multiple problem: First, to differentiate the several propositions which pass as the PLT by clarifying the sense of "private language" being used; second, to determine which ones are true; and third, to explain why they are supposed to be intimately related. These problems differ from the question, debated around 1930, of whether or not it is possible to start with a private language about one's sensations or "raw" feelings and arrive at the intersubjective and communicable language of science. (On this question, see Rudolf Carnap, "Psychology in Physical Language," and J. R. Weinberg, *An Examination of Logical Positivism*.)

### THE SENSE OF "IMPOSSIBLE"

In all the interpretations of the PLT, the word *impossible* is understood in a strong sense that is not easy to characterize precisely. Some philosophers speak of "logical impossibility," but they do not necessarily mean that private languages are impossible in the sense that unbounded triangular figures are impossible. The expression "unbounded triangular figure" reduces to the formal self-contradiction "*unbounded* figures *bounded* by three lines" by means of a substitution allowed by the definition of "triangle." But few philosophers would suggest that there is a similarly ready definition of "language" by means of which we can produce a formal self-contradiction "private so-and-so which is not private." The impossibility at issue is like (1) the impossibility of unextended red things (that is, the impossibility that something be red and yet lack

width or length), or (2) the impossibility of a cube with fewer than eight edges. These do not lead straightforwardly to formal contradictions, since there are no definitions for all the terms involved; they depend on implication relations that constitute the concepts involved in their statement. In the last analysis, the persistent rejection of (1) and (2) evidences the failure to understand the meanings of all the words involved, that is, the lack of some of the relevant concepts. But (1) is unprovable and obvious, and (2) only needs a trivial argument, while the PLT (if true) requires careful reasoning. We shall speak of *conceptual impossibility* to refer to any formal self-contradiction, to any impossibility which entails a formal self-contradiction, and to any a priori impossibility such as that found in the above examples (1) and (2).

### THE PRIVATE LANGUAGE THESIS

The most important propositions often discussed as the PLT, each embodying a different idea of private language, are the following:

PLT\*: It is impossible for a man to use a word with a meaning that nobody else could, even in principle, understand.

PLT-1: It is impossible for a man to use words that refer to private objects, that is, objects that nobody else could—even in principle—know. (For subtheses arising out of the ambiguities of “know,” see H.-N. Castañeda, “The Private-Language Argument.”)

PLT-2: It is impossible for a man who has always lived in isolation to possess a language, even if his sounds are understandable by another person.

Here the expressions “could not in principle” and “impossible” are meant to express conceptual impossibility. PLT\* allows that a man may use words with meanings that nobody else in fact understands, provided that they are understandable to other people in the appropriate circumstances. PLT-1 allows that a man may refer to objects that, in fact, he alone knows, but again others must be capable of knowing them in the appropriate circumstances. PLT-2 allows that a man, like Robinson Crusoe, keeps possession of a language he learned previously while living in a community of speakers.

Many philosophers assume that it is conceptually impossible for two persons to share one and the same immediate sensation. Many also hold that, in a strict sense of “know,” others do not really know whether one has a certain immediate sensation or not, precisely because they cannot share it. On these assumptions, a language about one’s own immediate sensations would be

a language of the sort that PLT-1 claims to be impossible. Indeed, such a language is customarily regarded as the would-be prototype of private language.

In general, on the assumption that (direct) knowledge of the referent of a word is required for understanding the meaning of the word in question, PLT\* entails PLT-1. On this assumption, a language about one’s own immediate sensations is also private in the way that PLT\* claims to be impossible.

PLT-1 does not entail PLT\*. A word might have a meaning understandable to only one person because the word itself is a private object in the sense of PLT-1, even though everybody may be acquainted with the physical objects it refers to. For example, the words of a person’s language might all be mental images of German written words, so that all his thinking would be a sort of mental reading of German. In this case, the referents of the words would be public, but the words themselves would be private and hence unintelligible to others.

PLT-2 neither entails nor is entailed by PLT-1. If PLT-2 is true, then if on the previous assumptions about sensations, one’s language about one’s own sensations is private in the sense of PLT-1, then one could still, in principle, invent such a language. Conversely, the truth of PLT-1 does not by itself make it impossible for an isolated person to invent a language about physical objects. Similarly, PLT-2 neither entails nor is entailed by PLT\*.

APPLICATIONS OF PRIVATE LANGUAGE THESES. The important claims made with the help of the PLT do require other assumptions, which in their turn play roles, as we shall see, in the defense of the PLT itself. The most natural and pervasive of these assumptions is the following:

(A) In the sense of “thinking” in which one can both have a false (or true) thought and draw inferences from what one thinks, it is conceptually impossible to think without possessing a language that is a means of thinking.

From this assumption and PLT-2, one can conclude that the fact that one thinks, guarantees the existence of other persons, namely, one’s fellow speakers of the same language. Thus, the solipsist who merely asserted that it is possible that he alone exists at the time he is thinking would be contradicting himself (an argument of this sort can be constructed with premises suggested by Rush Rhees in “Can There Be a Private Language?”). Of course, many philosophers have serious objections to (A).

The existence of hallucinations, illusions, and visual perspective leads many philosophers to characterize every

case of perception in terms of our apprehension of sense data or immediate impressions. Some have proceeded to espouse a phenomenalist program of “logical reconstruction” of physical objects and minds as systems of sense data; others, however, have subscribed to some form of realism, that is, the complete irreducibility of physical objects and minds to sense data. But all of them have recently been criticized on the ground that the language of sense data is private in either the sense of PLT\* or the sense of PLT-1. Here, in addition to (A), the critics need the following assumption:

(B) If it is conceptually impossible that there be a language about entities of a sort *T*, then there are no entities of sort *T*.

Again, some philosophers would claim against (B) that if PLT\* or PLT-1 is true, then sense data or the given in experience are simply ineffable.

Many philosophers have subscribed to some form or other of a principle of verification, for example:

(C) It is conceptually impossible to understand a sentence without knowing what state of affairs would verify the statement made with it.

Assumption (C) leads to the view that language about states of consciousness is private, if we add to it and (A) and (B) the following principle:

(P) Only the person himself can verify conclusively and directly that he has certain experiences.

On this view, for instance, when someone else speaking about me says, “He is in pain,” he cannot understand or mean exactly the same thing that I understand and mean when, of myself, I say, “I am in pain.” But if PLT\* is accepted, one is involved in a contradiction. Here many philosophers have given up (P), and in order to guarantee that everybody else can know what somebody is feeling or thinking, some philosophers have espoused some form of behaviorism, that is, a view according to which every description of a person’s experiences or mental states is really shorthand for (synonymous with) a description of his bodily movements, his relations to other bodies, and his abilities to perform further movement. This is often supplemented with the supposition that first-person utterances like “I have a headache” do not make statements of direct knowledge but are, rather, learned responses, analogous to the natural responses of moaning, crying, and so on, which are said to constitute the person’s ache. As is to be expected, other philosophers have preferred to keep (P) and reject one or more of the other premises, in particular (C) or PLT\*. (See Castañeda,

op. cit., Part B, for a discussion of the privacy of experiences.)

THE MAIN ARGUMENTS FOR THE PLT. There are many arguments seeking to prove that being private makes it impossible for a language to have a property required for the existence of a language. Most of the arguments depend on the following assumption:

(D) A language is a system of rules, and to speak or write a language is to follow rules.

On this assumption, it suffices to establish the PLT to show that a man (say, Privatus) cannot be following rules when he is using a private language (to be called Privatish). This is, in fact, what a series of arguments suggested by Wittgenstein purports to do. The gist of the argument is as follows: A rule is, by its very nature, the sort of thing that can be misapplied (or disobeyed), but Privatus cannot misapply the rules of Privatish; hence, when speaking Privatish, Privatus is not following rules. The specific arguments are meant to support the crucial premise:

(1) Privatus cannot misapply the rules of Privatish.

A fair objection to (1) is that Privatus can certainly make slips; he may call something of kind *A* “*B*,” whatever “*A*” and “*B*” may mean in Privatish. Slips of the tongue are precisely ways in which one violates the rules (if there are such) of natural languages. For instance, if there are rules of English governing the application of color words to physical objects; whenever one commits a slip of the tongue and calls a red object “blue,” then one misapplies either a rule governing the use of “red” or one governing the use of “blue.”

This reply to (1) is often met by several rejoinders. The first claims both that a slip counts as a misapplication of a linguistic rule only if there is a way in which the speaker can in principle detect and correct his slip and that Privatus cannot detect or correct his slips. This rejoinder, however, changes the issue, since premise (1) says nothing about verifying the existence of a misapplication of a rule. Nevertheless, the rejoinder has a point, for if to use words is to apply rules, then one must at least sometimes be able both to know of one’s misapplications of the rules for the use of one’s words and to know how to make the appropriate corrections. The question of whether or not Privatish allows this is discussed below under premise (2).

The second rejoinder is that to obey a rule is a *custom* (use, institution), but Privatus’s actions cannot constitute a custom (see Ludwig Wittgenstein, *Philosophical Investi-*

gations, Sec. 199). This rejoinder would establish PLT-2 but not PLT\* or PLT-1. For it may be a custom in a tribe that people use words which they alone understand in the ways required by PLT\* or PLT-1. But as an argument for PLT-2 the rejoinder is by itself question-begging. It must be supported by an argument which shows that obeying a rule is indeed a custom.

The third rejoinder is that Privatus's slips do not count as violations of the rules of Privatish because we cannot be corrected or taught by others what is the correct thing to say (see Wittgenstein, *op. cit.*, Sec. 378, and Norman Malcolm, "Discussion of Wittgenstein's *Philosophical Investigations*," pp. 536f.). If the "cannot" here is taken to mean conceptual impossibility, the rejoinder does not apply to PLT-2. If it is taken in a weaker sense, that is, a sense in which a person may be in the position of being in fact corrected by other persons, then the rejoinder supports PLT-2, but it would not allow that there be just one language-user in the universe. Besides, it is not clear that it would allow that Antonia Udina, for example, used language when, as we normally say, he spoke Dalmatian as the last speaker of Dalmatian. Although a person who uses words must be capable of self-correction, it is not immediately obvious that a person's sounds cannot count as utterances of words if nobody else can (in some sense) correct him. The need for others' possible corrections has to be established by an argument. Thus, we are again thrown back to the other lines of reasoning.

The fourth rejoinder is that Privatus's slips do not count because another person, by noting Privatus's behavior and circumstances, cannot discover that his use of the word is correct or incorrect (adopting Malcolm, *op. cit.*, p. 537). This rejoinder also leaves PLT-2 unsupported if "cannot" is understood as expressing conceptual impossibility. While it must be conceptually possible for Privatus to know whether his uses of language are correct or incorrect, it is not at all clear that it must be possible for others to know this fact. The principle that it must be possible for others to know whether his uses of language are correct or incorrect requires an independent argument to support it. However, the present rejoinder has a point. It reminds us that if there is no way at all of telling, for any word of Privatish, whether or not Privatus used it correctly (however coherent the concept of a private language is), it would be a completely gratuitous hypothesis that Privatus spoke a private language. Although our topic here is only the conceptual possibility of private language, we should note that the claim that somebody's entire language is of the type described in PLT\* is cer-

tainly gratuitous. Yet the claim that someone has a mixed language, part of which is private in the sense of PLT\*, does not seem gratuitous.

The fifth rejoinder dismisses mere slips on the ground that they show at most a breakdown of a linguistic habit. The rejoinder asks us to consider the case of Privatus trying deliberately to apply a rule of Privatish and failing to comply with it. The rejoinder claims that, for Privatish, "thinking one was obeying a rule would be the same thing as obeying it," but "to *think* one is obeying a rule is not to obey a rule. Hence it is not possible to obey a rule 'privately'" (Wittgenstein, *op. cit.*, Sec. 202). This rejoinder does not require that every utterance of a word be a case of deliberately attempting to obey the corresponding linguistic rule(s). Conjoined with assumption (A), this view would lead to a vicious infinite regress. For then, in order to say something, one would have to be aware of the rules governing the words one intends to utter, and these rules in their turn would be formulated in some words the rules governing which one would have to be aware of through some other words, and so on ad infinitum. Therefore, to use language is, of necessity, to use most of the words from habit, not in intended obedience of the linguistic rules. The rejoinder cannot even demand that Privatus sometimes be aware of the rules of Privatish: A being might speak a language without ever rising to the level of formulating any of his rules. But if, by assumption (D), languages are made up of rules, then if it were conceptually impossible for Privatus to be at least sometimes aware of the rules of Privatish, Privatish would be a very defective language indeed, incapable of discharging the philosophical duties that private languages are alleged to discharge. Thus, the rejoinder is right in urging that

(a) For every rule  $R$  of a language  $L$  and every speaker  $S$  of  $L$ , it is conceptually possible that sometimes  $R$  applies to  $S$ 's situation while  $S$  thinks that he is obeying  $R$  without  $S$ 's actually obeying  $R$ .

Presumably, a rule of language is here of the form "If  $x$  is  $\phi$ , you may (must) call it '...,'" but the meaning of "call" is difficult. In one normal sense of "call," slips of the tongue are, again, ways in which (a) is true. Clearly, a person may think that he is calling a thing "red" in deliberate compliance with the English rule for "red," without realizing that he actually called it "blue" because he is deaf or because he simply did not hear what he said. In the same sense of "call," (a) can be true because the speaker deliberately calls a red thing "blue," if he thinks that the rule in question allows (or prescribes) his calling it "blue." In particular, suppose that the rule  $R$  allowing (or prescrib-

ing) that one call a thing “red” is the rule Gaskon typed yesterday and that today, confusedly, Gaskon thinks that the rule he typed yesterday allows (or prescribes) that a certain thing be called “blue,” and he calls the thing in question “blue,” thinking that he is complying with the rule. Here, in spite of his deliberately calling a certain thing “blue,” Gaskon’s use of “blue” and the rule he thinks he is complying with both satisfy (a). Both ways of satisfying (a) are open to Privatus. It might be argued that Privatus’s deliberately calling one of his private objects “A” instead of “B” has no point or “function” (see Wittgenstein, *op. cit.*, Sec. 260), since he is not talking to others. This is, however, false. Privatus may very well play word games involving miscallings of things. But more importantly, whether or not there is a point in Privatus’s flouting of the rules of Privatish has nothing to do with the issue about the possibility of private language.

The rejoinder often uses a stronger sense of “call.” In this sense, by a natural development of assumption (A), to *think* that something is, for example, red is to call it “red.” This stronger sense appears in an argument given in support of PLT-1. As said above, language about one’s own immediate sensations is often regarded as the paradigm of private language in the sense of PLT-1. Now, one knows incorrigibly that one’s sensations have immediately sensible qualities. That is to say, if one believes that one has a pain (itch, tickling, feeling of discomfort), then one knows that one has a pain (itch, tickling, feeling of discomfort). So it is impossible to have no pain while one thinks that one has a pain. Thus, if one thinks that one is obeying the rule of the form “If  $x$  is a pain, you may (must) call it ‘pain,’” one surely thinks that one is in pain and the rule cannot fail to apply. Similarly, since one also has incorrigible knowledge of the absence of one’s immediate sensations, if the objects that Privatus can think about in Privatish are only his immediate sensations, then when he thinks that a rule of Privatish does not apply, the rule does not, in fact, apply. But if “call” is taken in its normal sense, neither of these two features of the rules of Privatish implies that Privatus cannot think that he is obeying a rule (which then applies) without actually obeying it, since slips and deliberate miscallings are still available as violations of the rule. However, if “call” is taken in the strong sense (in which thinking can be calling), then if Privatus thinks that he is obeying a rule of the form “If  $x$  is  $A$ , you may (must) call it ‘A,’” he surely thinks that the rule applies, that is, he thinks that the object  $x$  is  $A$ ; if  $A$  is a sensible property of Privatus’s immediate sensation  $x$ , then  $x$  is  $A$ , and Privatus is both calling  $x$  “A” and unavoidably obeying the rule. Thus, if Privatish is a private language about Privatus’s immediate

sensations and their sensible properties, then (a) above and (b) below are both false:

(b) For every rule  $R$  of a language  $L$  and every speaker  $S$  of  $L$ , it is possible that sometimes  $S$  thinks that he is obeying  $R$  while he is not.

Since (a) is true, Privatish is not a private language.

This argument does not by itself support PLT-2; it may or may not support PLT\*, depending on how one interprets the phrase “knowing the meaning of a word.”

There is, however, a difficulty with the above argument. Consider the rule of English: “If  $x$  is a cat, you may (must) call  $x$  ‘cat’; that is, you may (must) think that  $x$  is a cat.” This rule differs from the above rule for the Privatish word “A” in that thinking that one is obeying the rule for “cat” does not imply that the rule for “cat” applies to the situation in question. For to think that one is obeying the latter rule implies that one thinks that it applies, and this implies that one thinks that some object  $x$  is a cat. But surely one can be mistaken about  $x$ ’s being a cat. Yet the rule for “cat” also fails to satisfy condition (a). Suppose that the rule applies; then the object  $x$  in question is a cat. And suppose that one thinks that one is obeying the rule; then it is true that one thinks that if  $x$  is a cat one may (must) think that  $x$  is a cat, and that one thinks that  $x$  is cat. Thus, one is in fact obeying the rule! Therefore, the strong sense of “call” included in the concept of language rule  $R$  makes (a) an impossible condition.

Now, in the case in which a rule  $R$  does not apply to a man’s situation, we are often reluctant to say that when such a man thinks that he is obeying  $R$ , he is not obeying  $R$ . But we could say this with no great distortion, and if we did, we could say that the above rule for the English word “cat” satisfies condition (b). For in a situation in which an object  $x$  is not a cat and the rule does not apply, we may very well both misperceive or otherwise think that  $x$  is a cat and think that, in accordance with the rule, we may (must) think that  $x$  is a cat. Thus, if we raise (b) as the crucial condition that linguistic rules must satisfy, then we can claim that PLT-1 is established in the sense that a pure language of sensations is impossible. But this answer is inconclusive. Besides the small amount of distortion involved, there is the fact that (b) is not a general condition of rules. This is shown by the following rule which a man might give to his son: “If you think that you need to delay your action, think that  $1 + 2 + 3 + \dots + 24 = 300$ .” Since to think that one thinks that  $p$  entails that one thinks that  $p$ , if the boy thinks that the rule applies, he thinks he needs to delay his action, and the rule applies. If he thinks that he is obeying the rule, he thinks

both that it applies and that  $1 + 2 + 3 + \dots + 24 = 300$ ; hence he thinks that  $1 + 2 + \dots + 24 = 300$ ; hence, the rule applies and he obeys it. Thus, to defend PLT-1 by means of (b) requires an independent argument showing that rules of language must, in any case, comply with (b), distorted as suggested.

Let us turn now to a subtler line of argument. Some defenders of the PLT do not argue for (1) but for

(2) Privatus cannot distinguish his correct uses of Privatish words from his incorrect uses.

Suppose, then, that Privatus is debating whether something is *A* or not. Suppose that Privatish is private in the sense of PLT-1. Here the defenders of the PLT adduce (a) that Privatus lacks a *criterion* of correctness, that is, “something *independent* of his impression” that he is correctly using the Privatish rule governing the use of “*A*” by means of which he can “prove his impression correct” (Malcolm, *op. cit.*, p. 532), and (b) that his impression that he remembers what objects of kind *A* appeared like before is of no help, since memory “is not the highest court of appeal” (Wittgenstein, Sec. 56) and the “process [of checking memories] has got to produce a memory which is actually *correct*” (Sec. 265). Now, these points exaggerate Privatus’s predicament. Privatus’s private objects may be related among themselves by entailment, by coexistence, by similarities, by causal relationships, and so on. Privatus can resort to any of these to test whether he is, on the present occasion, using the term “*A*” correctly. For instance, in Privatish, “being *A*” may be logically equivalent to “being *B* and becoming *C* in the presence of another *C*.” Indeed, Privatus may even employ paradigms. The very first object he calls “*A*” may very well be enduring, so that he can compare the next objects of kind *A* with it. The same applies to languages of the type mentioned in PLT-2. Furthermore, memory is the highest court of appeal when it comes to our knowledge of the past. True, we have records and other historical evidence, but all of this only provides inductive evidence, not a proof, and our inductions involve the acceptance of unchallenged memories.

Nevertheless, Privatus is not only in no position to question the correctness of all of his uses of words, but he also cannot *prove* that the uses he questions are correct unless he is allowed the ability to identify certain properties of objects without criteria and without challenging his memory. But exactly the same happens with the speakers of *any* language. In the case of terms like “red” and “straight,” for instance, there is nothing at all to which an English speaker *E* can resort in order to “prove” that he has correctly called an object red or straight. His

fellow speakers may all utter in unison, “Not red but blue.” Yet this choral utterance is not a proof; the speakers may be lying, may all be victims of a hallucination, or may just be rehearsing a new song—or the whole proceedings may be just *E*’s hallucination. In any case, for *E* to accept the correction, he must correctly identify the words expressing it without the use of criteria and remember correctly the meanings of these words. A vicious infinite regress would ensue if *E* were required to have a proof that he both remembers this correctly and identifies the objects the words apply to.

Moreover, there is nothing to prove each corrector’s use of words correct. Suppose, for example, that one corrector learned the meaning of “blue” with the help of object *O* and that he continuously stares at *O* during the preceding two minutes before correcting Privatus. He still must *remember* correctly that *O* has the same color it had two minutes before, that the color of *O* is called “blue,” that the name of the color sounds “b–l–u–e,” that the noise “red” uttered by *E* has the same meaning that makes red and blue incompatible, and so on. Thus, either somebody *just* identifies some words or objects correctly and remembers some qualities of objects and the meanings of some words correctly, or else nobody can be corrected by another speaker. In sum, demands (a) and (b) cannot be adduced against the possibility of a private language.

**LOGICAL WORDS.** Often it is claimed that a private language cannot have logical words or syntactical rules, both of which are necessary for the existence of logical relationships. Clearly, if a private language is allowed no implications or entailments, it would certainly be no language. But if “private language” is meant in the sense of PLT\* or PLT-1 or PLT-2, this contention appears to be false. Often this contention is defended on the ground that a really private language does not have words with meanings in common with the words of another language (Wittgenstein Sec. 261; Malcolm p. 537). Now, private language in this sense is impossible. A language is a system of words of which some refer to objects, some signify properties or relations, and some express logical connections; the words expressing logical connections must be capable of being understood by anybody else and must, therefore, be common to all languages. This is an important result. But it is not the same as PLT\*, which requires that every single word of a language must be understood by persons other than the speaker. Likewise, the impossibility of languages without logical words does not imply that a language cannot have some nonlogical words which refer to private objects, that is, it does not imply that PLT-1 is true. Again, that a language must have

logical words implies nothing about the possibility of a single man developing a language for and by himself, that is, does not imply that PLT-2 is true.

“THE SAME.” Apparently Wittgenstein knew that there are no criteria (in the sense of something independent) which prove that words have been used incorrectly. He also knew that the correctness of an application of a word is not determined by a rule whose formulation serves as a recipe or canon. His fundamental opposition to private language derives from his profound investigations into the nature of concepts and his strong inclination toward an extreme nominalism. This opposition is never crystallized in a definite argument, but its gist is, in crude form, as follows. Postulate:

(E) The similarities and samenesses we find in things do not exist *in rerum natura*, that is, do not exist in things as we find them, independently of our finding them or of our referring to them in the way we do; they “come from the language” (Rhees p. 80) and at bottom consist of the fact that we “call” the things in question the same (Wittgenstein Secs. 146, 149, 185–190, 208–223, 348–352).

On a rigorous interpretation of (E), we find a rationale for assumptions (A), (B), and (C), as well as for the fact that the PLT has a chameleon-like and pervasive character. If we take (E) literally, then to find a property in several things is to find that we “call” the things in question “the same” or refer to them with the same word. Thus, it is impossible to think that something is such-and-such without a language in which there is an expression (even if a very long phrase) which “constitutes” the such-and-such in question. This is assumption (A). Also, (B), without an expression “constituting” a type *T*, there is no type *T* for things to belong to. Similarly, to understand an expression is not to apprehend an independently existing (or subsisting) property but simply to know how and to what to apply it, and this includes knowing how to call certain utterances “true” in which the expression is correctly applied. This is, in fact, a generalization of assumption (C).

We cannot say that a man in doubt about whether or not he used a word correctly must simply identify certain features of things without criteria and, armed with these identifications, test his uses of words. For on the extreme interpretation of assumptions (A) through (E) to identify a feature is to “call” a thing something. So, when the use of a word is at issue, the identification and nature of the thing is precisely what is at issue. The referents of one’s previous uses of the word, as well as the uses themselves, are irrelevant. If one “calls” something “A,” then it is A

and *a fortiori* similar to the previous A’s; if one withholds the name “A” from it, then it is not an A, and a fortiori it is dissimilar to all A’s with respect to being an A. Clearly, it does not matter whether one’s language is about private or about public objects; one’s uses of words simply fail to be capable of being incorrect. They would seize reality so well that each “would have to be at once a statement and a definition” (Rhees p. 82).

Thus, the following question arises. If, on assumption (D), language is a matter of rules and rules are the sort of thing that can be misapplied or not, how, then, is language possible after all? At this stage, obviously, we are not interested in proving anything but are anxious to find an explanation. Wittgenstein seems to suggest one: A man’s uses of words can be incorrect only if they are compared with those of his fellow speakers. His “calling” something “A” is correct if his cospeakers now also call it “A.” Then it is A and a fortiori similar to the things he and his cospeakers previously called “A.” That is why obeying a rule of language is a practice (Wittgenstein, op. cit., Sec. 202). It is not necessary that the speakers of the language should call the thing in question “A” or that they call it “A” afterward. Nor is it necessary that they call it “A” or anything at all, or that they call it the same thing. It is just a contingent fact that they coincide in calling it “A.” But this coincidence (or agreement) is an empirical fact that is necessary for the existence of language.

Such is the underlying argument of Wittgenstein’s remarks (Secs. 146, 149, 185–190, 208–223; for a discussion of the role of Wittgenstein’s extreme nominalism in his views about necessary truth, see Michael Dummett’s “Wittgenstein’s Philosophy of Mathematics”). He builds a Heraclitean picture of language as something living only in our actual use of it and changing according to our needs. But is this a true picture of the connection between language and reality?

Here we cannot discuss the whole issue of nominalism, but to this writer it seems indefensible. We could doubtless have classified objects in entirely different ways from the ways we in fact do. For instance, we might have had no color words, no terms for species of plants or animals, and instead have used, say, “sha” for some elephants and white roses and reddish sand, and “sho” for female elephants, eggs, and rivers. But even so, we should have had to *find* features of similarity in the things so classified, and these features would have provided tests for the correct application of our words. At any rate, the view that things are the same because we “call” them “the same” or because we refer to them with the same words can get off the ground only by postulating our recogni-



tion of the samenesses of words, that is, the similarities of noises whose application to things constitutes the similarities of the latter. A serious infinite regress would ensue if we also hold that our words are similar only because we “call” them so.

The several propositions that are often debated as the claim that private languages are impossible can be linked to each other only under the assumption of extreme nominalism. None of the arguments given for the claim appear to be successful. There may be no conclusive way of either proving or refuting this claim. Perhaps the only course is to build detailed and rigorous philosophical views on each alternative and assess the adequacy of such views by their consequences. This topic continues to be widely discussed in the literature, and many philosophers adopt a position different from that advocated in the present article.

**See also** Carnap, Rudolf; Malcolm, Norman; Mind-Body Problem; Rule Following; Solipsism; Wittgenstein, Ludwig Josef Johann.

## Bibliography

### THE PROBLEM BEFORE WITTGENSTEIN

Carnap, Rudolf. “Psychology in Physical Language.” In *Logical Positivism*, edited by A. J. Ayer. Glencoe, IL: Free Press, 1959. Ch. 8. Translated by Frederic (there called George) Schick from the German version, which appeared in *Erkenntnis* 3 (1932–1933).

Weinberg, J. R. *An Examination of Logical Positivism*. New York and London, 1936. Ch. 11.

### WITTGENSTEIN’S VIEWS

Albritton, Rogers. “On Wittgenstein’s Use of the Term ‘Criterion.’” *Journal of Philosophy* 56 (1959): 845–857. An excellent collation and exegesis of the passages in which Wittgenstein discusses criteria; brings out certain internal tensions in Wittgenstein’s conception.

Dummett, Michael. “Wittgenstein’s Philosophy of Mathematics.” *Philosophical Review* 68 (1959): 324–348.

Malcolm, Norman. “Wittgenstein’s *Philosophical Investigations*.” *Philosophical Review* 63 (1954): 530–559. An extended review and discussion that presents Wittgenstein’s arguments against the PLT, with emphasis on the private language speaker’s lack of a criterion of correctness.

Rhees, Rush. “Can There Be a Private Language?” *PAS*, Supp., 28 (1954): 77–94. A symposium with A. J. Ayer containing an exposition of Wittgenstein’s views on private language understood mainly as PLT-2; conveys Wittgenstein’s nominalistic rationale very well, without discussing it as such.

Wittgenstein, Ludwig. *Philosophical Investigations*. German with facing translation by G. E. M. Anscombe. Oxford: Blackwell, 1953.

### THE PROBLEM AFTER WITTGENSTEIN

Ayer, A. J. “Can There Be a Private Language?” *PAS*, Supp., 28 (1954): 63–76. A symposium with Rush Rhees in which Ayer criticizes the PLT and emphasizes the need for criterionless identification at some point.

Carney, James. “Private Language.” *Mind* 64 (1960): 560–565. Favors the PLT.

Castañeda, Héctor-Neri. “Criteria, Analogy, and Knowledge of Other Minds.” *Journal of Philosophy* 59 (1962): 533–546. Discusses some problems of Malcolm’s views on criteria.

Castañeda, Héctor-Neri. “The Private-Language Argument.” In *Knowledge and Experience*, edited by C. D. Rollins. 1962 Oberlin Philosophy Colloquium. Pittsburgh: University of Pittsburgh Press, 1963. A symposium with V. C. Chappell and J. F. Thomson; surveys critically the arguments for the PLT and also replies to the other symposiasts.

Chappell, V. C. “Comments.” In *Knowledge and Experience*, edited by C. D. Rollins. Vigorously defends the PLT from criticisms raised by Castañeda’s discussion.

Garver, Newton. “Wittgenstein on Private Language.” *Philosophy and Phenomenological Research* 20 (1960): 389–396. Favors the PLT.

Hardin, Clyde L. “Wittgenstein on Private Language.” *Journal of Philosophy* 56 (1959): 517–528. Attacks the PLT.

Hervey, Helen. “The Private Language Problem.” *Philosophical Quarterly* 7 (1957): 63–79. Accepts PLT\* but claims that the real problem is whether or not one can recognize and classify one’s sensations apart from their outward expression.

Tanburn, N. P. “Private Languages Again.” *Mind* 72 (1963): 88–102. Attacks PLT-2 and the thesis that all entities referred to in conversation must be publicly available.

Thomson, J. F. “Comments.” In *Knowledge and Experience*, edited by C. D. Rollins. In his part of the symposium with Castañeda and Chappell, Thomson contends that the PLT is too unclear for decision.

Todd, W. “Private Languages.” *Philosophical Quarterly* 12 (1962): 206–217. Argues that “a private language, in the sense of a personal sensation language, is logically possible,” refutes certain arguments of Malcolm, and attempts to show that a private language need not lack criteria of correctness.

Wellman, Carl. “Wittgenstein and the Egocentric Predicament.” *Mind* 68 (1959): 223–233. Discusses critically some of Wittgenstein’s arguments for the PLT in connection with the view that all knowledge is based on private experiences.

*Héctor-Neri Castañeda (1967)*

## PRIVATE LANGUAGE PROBLEM [ADDENDUM]

Although the proper formulation and assessment of Ludwig Wittgenstein’s argument (or arguments) against the possibility of a private language continues to be disputed, the issue has lost none of its urgency. At stake is a broadly

Cartesian conception of experiences that is found today in much philosophy of mind.

### WHAT IS A PRIVATE LANGUAGE?

In §243 of *Philosophical Investigations* (1967; see also §256) Wittgenstein introduces the idea of a language in which “a person could write down or give vocal expression to his inner experiences—his feelings, moods, and the rest—for his private use. ... The individual words of this language are to refer to what can only be known to the person speaking; to his immediate private sensations. So another person cannot understand the language.” In subsequent sections (according to some commentators, up to as far as §315) Wittgenstein criticizes the possibility of such a “private language,” and this is where “the private language argument” is usually supposed to be located.

Following the main essay, suppose that Privatus speaks Privatish, a private language. §243 suggests that Privatish has two features:

- (1) Privatish contains a referring expression *n* that refers to one of Privatus’s “immediate private sensations” *s*.
- (2) *s* “can only be known to” Privatus.

According to Wittgenstein Privatish has a third feature, which he apparently thinks follows from (1) and (2):

- (3) *n* can only be understood by Privatus (and, hence, Privatish can only be completely understood by Privatus).

At this point three questions arise. First, what does (2) mean? Second, what sort of referring expression is *n*? Third, why is (3) supposed to follow from (1) and (2)?

By (2) Wittgenstein seems to mean that only Privatus can know whether he is having *s*. “The essential thing about private experience is ... that nobody knows whether other people also have *this* or something else” (§272, see also §246). Of course, this conception of sensations is held by Wittgenstein’s opponent (a defender of the possibility of a private language), not Wittgenstein himself.

As to the referring expression *n*, it is a name, not a description (e.g., “the private sensation caused by pin-pricks”) (see §§256–257). Not even Wittgenstein’s opponent would accept that to understand a description that in fact refers to sensation *s* one has to know that it refers to *s*.

Is *n* a proper name of a token sensation, or is it a common noun referring to a type of sensation? If *n* refers

to a token sensation, something occurring only in the mind of Privatus, then Wittgenstein’s opponent looks exactly like Bertrand Russell in “The Philosophy of Logical Atomism” (1918/1956; see also Candlish 2004). According to Russell, “[i]n order to understand a proposition in which the name of a particular occurs, you must already be acquainted with that particular” (1918/1956, p. 204; see also Russell 1912, chapter 5). Since, on Russell’s view, the only particular things with which one is acquainted are private items he calls *sense-data* (and, in addition, perhaps one’s self), no two people can be acquainted with the same particular, and so no two people can understand a genuine name (as Russell puts it, a name in the “logical sense”). Hence, Russell thinks, if Privatus’s name *n* refers to a token sensation, no one else can understand it.

However, it is clear that Wittgenstein takes the sensations in question to be types, not tokens (see, in particular, §258); accordingly, the name *n* is a common noun. But then Russell’s views about acquaintance and understanding play no role in the justification of (3), for Russell holds that two people can be acquainted with the same property (or type), including properties of private objects. Thus, if Privatus is acquainted with a certain type of sensation *s*, that is no barrier, on Russell’s view, to others also being acquainted with *s*.

So why does Wittgenstein think that (3) follows from (1) and (2)? His argument may be this: one cannot know that Privatus’s name *n* refers to *s*, so one cannot know what *n* means, and hence one cannot understand it. But it is not obvious that knowledge that *n* refers to *s* is necessary for understanding *n*, or for successful communication using it: perhaps all that is required is that one believes that Privatus’s name *n* refers to *s*. The upshot is that Wittgenstein’s double characterization of a private language as one “which describes my inner experiences and which only I myself can understand” (§256) is contentious. (This point is due to Edward Craig [1982]; for further discussion see Craig [1997].)

Given that the two characterizations of a private language should be separated, it is probably better to use the first, leaving the second as a disputed consequence. Thus, a private language may be explained as one containing names for types of inner experiences, with the further stipulation that, if there are any inner experiences, no one knows whether others have the same types of inner experiences as him- or herself.

§258

§258 contains the famous example of keeping “a diary about the recurrence of a certain sensation,” and most commentators identify it as the core of the private language argument. The key move in this section is to cast doubt on whether the diary keeper can “impress on [himself] the connexion between the sign [“S”] and the sensation,” and so “remember the connexion *right* in the future.” Anthony Kenny points out, against some commentators, that “remembering the connexion right” does not mean that one correctly applies “S” to one’s sensation, but that one remembers the meaning of S (1973, pp. 191–193).

Supposedly, there is no fact about the meaning of “S” for the diarist to remember because there is “no criterion of correctness.” Here, there is little consensus on what the missing criterion amounts to, or whether its absence does indeed show that the diarist fails to attach a meaning to “S.” For some representative examples of exegesis, see Malcolm Budd (1989, chapter 3), Stewart Candlish (1980, 2003), John V. Canfield (1991, 2001), Robert J. Fogelin (1987, chapter 12), P. M. S. Hacker (1986, chapter 9; 1990, 61–67); Colin McGinn (1997, chapter 4), David Pears (1988, chapters 13, 14, 15), Scott Soames (2003, chapter 2), and Crispin Wright (1986).

### KRIPKE’S WITTGENSTEIN

In *Wittgenstein on Rules and Private Language* (1982), Saul A. Kripke suggests (without unreservedly endorsing) a novel and exciting interpretation of the private language argument (see also Fogelin 1987, p. 241, n. 10). On this interpretation the main argument appears in the earlier long discussion of following a rule starting around §138. As Kripke observes, the conclusion of the private language argument is stated in §202, well before the argument’s traditional location, “Hence it is not possible to obey a rule ‘privately’: otherwise thinking one was obeying a rule would be the same thing as obeying it.” Kripke takes the sections following §243 to be a discussion of a purported counterexample—namely, sensation language—to the conclusion argued for earlier.

The argument Kripke extracts from Wittgenstein is in two parts. The first part purports to establish that there are no facts that make it the case that an individual (Jones, say) means something by a word (addition by “+,” say). This conclusion is reached by canvassing all the plausible candidates for such meaning-constituting facts and finding them all wanting. A *skeptical paradox* (“our paradox” of §201) looms, “no facts, no truth conditions, correspond to such statements such as “Jones means

addition by ‘+’” (1982, p. 77). The second part offers a skeptical solution: “skeptical” because the paradoxical conclusion is embraced; a “solution” because sentences such as “Jones means addition by ‘+’” remain assertible, despite the lack of any “corresponding fact.” And the account of why such sentences are assertible essentially involves a linguistic community, so that if Jones is “considered in isolation,” he cannot be said to mean anything by his words. This is the most general sense in which a “private language” is impossible: An individual considered in isolation from other speakers cannot be said to speak a language (see Kripke 1982, pp. 109–110).

Most commentators have not endorsed Kripke’s interpretation (in particular, see Baker and Hacker 1984, chapter 1; McGinn 1984, chapter 2). However, Kripke’s Wittgenstein has become a philosopher of considerable interest in his own right.

### THE COMMUNITY VIEW

Kripke’s book revived interest in the issue of whether the private language argument and related material on rule following is supposed to exclude a Robinson Crusoe isolated from birth from speaking a language (discussion of this topic goes back to Alfred J. Ayer [1954] and Rush Rhees [1954]; see also Kripke [1982, p. 110, n. 84]). While the characterization of a private language in §243 seems to leave room for such a Crusoe, other sections, notably §198, suggest the opposite. Norman Malcolm (1986, 1989) offers a defense of the “community view,” and is countered by G. P. Baker and Hacker (1990). The community view is rejected by most commentators; for further discussion and references, see Canfield (1996).

*See also* Rule Following.

### Bibliography

- Ayer, Alfred J. “Can There Be a Private Language?” In *The Philosophy of Wittgenstein*. Vol. 9, edited by John V. Canfield. New York: Garland, 1986. Ayer’s essay originally published in the journal *Proceedings of the Aristotelian Society* in 1954.
- Baker, G. P., and P. Hacker. “Malcolm on Language and Rules.” *Philosophy* 65 (1990): 167–179.
- Baker, G. P., and P. M. S. Hacker. *Scepticism, Rules, and Language*. Oxford, U.K.: Basil Blackwell, 1984.
- Budd, Malcolm. *Wittgenstein’s Philosophy of Psychology*. London: Routledge, 1989.
- Candlish, Stewart. “Private Language.” *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford, CA: Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University, 2003. <http://plato.stanford.edu/archives/spr2004/entries/private-language/>

- Candlish, Stewart. "The Real Private Language Argument." *Philosophy* 55 (1980): 85–94.
- Canfield, John V. "The Community View." *Philosophical Review* 105 (1996): 469–488.
- Canfield, John V., ed. *The Philosophy of Wittgenstein, Vol. 9: The Private Language Argument*. New York: Garland, 1986.
- Canfield, John V. "Private Language." In *Wittgenstein's Philosophical Investigations: Text and Context*, edited by Robert L. Arrington and Hans-Johann Glock. London: Routledge, 1991.
- Canfield, John V. "Private Language: The Diary Case." *Australasian Journal of Philosophy* 79 (2001): 377–394.
- Craig, Edward. "Meaning and Privacy." *A Companion to the Philosophy of Language*, edited by Robert Hale and Crispin Wright, 127–145. Oxford, U.K.: Blackwell, 1997.
- Craig, Edward. "Meaning, Use, and Privacy." *Mind* 91 (1982): 541–564.
- Fogelin, Robert J. *Wittgenstein*. 2nd ed. London: Routledge and Kegan Paul, 1987.
- Hacker, P. M. S. *Insight and Illusion: Themes in the Philosophy of Wittgenstein*. Rev. ed. New York: Oxford University Press, 1986.
- Hacker, P. M. S. *Wittgenstein: Meaning and Mind*. Oxford, U.K.: Basil Blackwell, 1990.
- Kenny, Anthony. *Wittgenstein*. London: Penguin, 1973.
- Kripke, Saul A. *Wittgenstein on Rules and Private Language: An Elementary Exposition*. Oxford, U.K.: Blackwell, 1982.
- Malcolm, Norman. *Nothing Is Hidden*. Oxford, U.K.: Blackwell, 1986.
- Malcolm, Norman. "Wittgenstein on Language and Rules." *Philosophy* 64 (1989): 5–28.
- McGinn, Colin. *Wittgenstein on Meaning*. Oxford, U.K.: Basil Blackwell, 1984.
- McGinn, Marie. *Routledge Philosophy Guidebook to Wittgenstein and the Philosophical Investigations*. London: Routledge, 1997.
- Pears, David. *The False Prison*. Vol. 2. Oxford, U.K.: Clarendon Press, 1988.
- Rhees, Rush. "Can There Be a Private Language?" In *The Philosophy of Wittgenstein*, edited by John V. Canfield. New York: Garland, 1986. Rhees's essay originally published in the journal *Proceedings of the Aristotelian Society* in 1954.
- Russell, Bertrand. "The Philosophy of Logical Atomism" (1918). In *Logic and Knowledge: Essays, 1901–1950*, edited by Robert Charles Marsh. London: George Allen and Unwin, 1956.
- Russell, Bertrand. *The Problems of Philosophy*. New York: Henry Holt, 1912.
- Soames, Scott. *Philosophical Analysis in the Twentieth Century*. Vol. 2. Princeton, NJ: Princeton University Press, 2003.
- Wittgenstein, Ludwig. *Philosophical Investigations*. 2nd ed. Translated by G. E. M. Anscombe. Oxford, U.K.: Blackwell, 1967.
- Wright, Crispin. "Does *Philosophical Investigations I*. 258–60 Suggest a Cogent Argument against Private Language?" In *Subject, Thought, and Context*, edited by Philip Pettit and John McDowell, 209–266. New York: Oxford University Press, 1986.

Alex Byrne (2005)

## PROBABILITY AND CHANCE

The weather report says that the chance of a hurricane arriving later today is 90 percent. Forewarned is forearmed: Expecting a hurricane, before leaving home I pack my hurricane lantern.

Probability enters into this scenario twice, first in the form of a physical probability, sometimes called a chance, quantifying certain aspects of the local weather that make a hurricane very likely, and second in the form of an epistemic probability capturing a certain attitude to the proposition that a hurricane will strike, in this case one of considerable confidence.

It is not immediately obvious that these two probabilities are two different kinds of thing, but a *prima facie* case can be made for their distinctness by observing that they can vary independently of one another: For example, if the meteorologists are mistaken, the chance of a hurricane may be very low though both they and I am confident that one is on its way.

Most philosophers now believe that the apparent distinctness is real. They are therefore also inclined to say that my belief that the physical probability of a hurricane is very high is distinct from my high epistemic probability for a hurricane. There must be some principle of inference that takes me from one to the other, a principle that dictates the epistemic impact of the physical probabilities—or at least, of my beliefs about the physical probabilities—telling me, in the usual cases, to expect what is physically probable and not what is physically improbable. One can call such a principle, mediating as it does between two different kinds of probability, a probability coordination principle.

The three principal topics of this entry will be, in the order considered, epistemic probability, physical probability, and probability coordination. Two preliminary sections will discuss the common mathematical basis of epistemic and physical probability and the classical notion of probability.

### THE MATHEMATICAL BASIS

What all probabilities, epistemic and physical, have in common is a certain mathematical structure. The most important elements of this structure are contained in the axioms of probability, which may be paraphrased as follows:

- (1) All probabilities are real numbers between zero and one inclusive (for any proposition  $a$ ,  $0 \leq P(a) \leq 1$ ).
- (2) The probability of an inconsistent proposition is zero; the probability of a logical truth, or tautology, is one.
- (3) The probability that either one or the other of two mutually exclusive propositions is true is equal to the sum of the probabilities of the individual propositions. (Two propositions are mutually exclusive if they cannot both be true; the *cannot* is interpreted as a matter of logical consistency, so that the axiom says that for any two propositions  $a$  and  $b$  such that  $a \vdash \neg b$ ,  $P(a \vee b) = P(a) + P(b)$ .)

The axioms as stated here assume that probabilities are attached to propositions, such as the proposition that “A hurricane will strike New York at some time on the afternoon of January 20, 2005.” The axioms may also be stated in a way that assumes that probabilities attach to events. It is more natural to attach epistemic probabilities to propositions and physical probabilities to events, but when the two kinds of probability are discussed side by side it is less confusing, and quite tolerable, to take propositions as the primary bearers of both kinds of probability. Nothing important is thought to turn on the choice.

The three axioms of probability, though simple, may be used to prove a wide range of interesting and strong mathematical theorems. Because all probabilities conform to the axioms, all probabilities conform to the theorems. It is possible, then, to do significant work on probability without presupposing either epistemic or physical probability as the subject matter, let alone some particular construal of either variety. Such work is for the most part the province of mathematicians.

Philosophical work on probability may also be mathematical, but is most often directed to one or the other variety of probability, usually attempting a philosophical analysis of probability statements made in a certain vein, for example, of probability claims made in quantum mechanics or evolutionary biology (both apparently claims about physical probability) or of probability claims made in statistical testing or decision theory (both apparently claims about epistemic probability).

Two important notions encountered in statements of the mathematical behavior of probability are conditional probability and probabilistic independence. Both are introduced into the mathematics of probability by way of definitions, not additional axioms, so neither adds anything to the content of the mathematics.

**CONDITIONAL PROBABILITY.** The probability of a proposition  $a$  conditional on another proposition  $b$ , written  $P(a|b)$ , is defined to be  $P(ab)/P(b)$ , where  $ab$  is the conjunction of  $a$  and  $b$ . (The conditional probability is undefined when the probability of  $b$  is zero.) For example, the probability of obtaining three heads on three successive tosses of a coin, conditional on the first toss yielding heads, is the probability of obtaining three heads in a row, namely one-eighth, divided by the probability of obtaining heads on the first coin, namely one-half—in other words, one-quarter.

Some writers suggest taking conditional probability as the basis for all of probability mathematics, a move that allows, among other things, the possibility of conditional probabilities that are well defined even when the probabilities of the propositions conditionalized on are zero (Hájek 2003). On this view, the mathematical posit stated above linking conditional and unconditional probabilities is reinterpreted as an additional axiom.

The act of conditionalization may be used to create an entirely new probability distribution. Given an old probability distribution  $P(\dot{s})$  and a proposition  $b$ , the function  $P(\dot{s}|b)$  is provably also, mathematically speaking, a probability distribution. If  $k$  is a proposition stating all of one’s background knowledge, for example, then a new probability distribution  $P(\dot{s}|k)$  can be formed by conditionalizing on this background knowledge, a distribution that gives, intuitively, the probabilities for everything once one’s background knowledge is taken into account. This fact is especially important in the context of epistemic probability.

**PROBABILISTIC INDEPENDENCE.** Two propositions  $a$  and  $b$  are probabilistically independent just in case  $P(ab) = P(a)P(b)$ . When the probability of  $b$  is nonzero, this is equivalent to  $P(a|b) = P(a)$ , or in intuitive terms, the claim that the truth or otherwise of  $b$  has no impact on the probability of  $a$ .

Several of the most important and powerful theorems in probability mathematics make independence assumptions. The theorem of most use to philosophers is the law of large numbers. The theorem says, very roughly, when a large, finite set of propositions are independent, but have the same probability  $p$ , then the proportion of propositions that turn out to be true will, with high probability, be approximately equal to  $p$ . (The generalization to countably infinite sets of propositions is easy if the propositions are ordered; substitute *limiting frequency* for *proportion*.)

For example, the propositions might all be of the form “Coin toss  $x$  will produce heads,” where the  $x$  stands for any one of a number of different tosses of the same coin. If the probability of each of the propositions is one-half, then the law of large numbers says, in effect, that provided the tosses are independent, it is very likely that about one-half will yield heads.

It is natural to interpret the probabilities in this example as physical probabilities, but the law of large numbers applies equally to any kind of probability, provided that independence holds. There are, in fact, many variants of the law of large numbers, but the details are beyond the scope of this entry.

### CLASSICAL PROBABILITY

The development of the mathematics, and then the philosophy, of probability was spurred to a perhaps surprising degree by an interest, both practical and theoretical, in the properties of simple gambling devices such as rolled dice, tossed coins, and shuffled cards. Though there was from the beginning a great enthusiasm for extending the dominion of the “empire of chance” to the ends of the earth, gambling devices were—and to some extent are still—the paradigmatic chance setups.

A striking feature of gambling devices is their probabilistic transparency: The discerning eye can “read off” their outcomes’ physical probabilities from various physical symmetries of the device itself, seeing in the bilateral symmetry of the tossed coin a probability of one-half each for heads and tails, or in the six-way symmetry of the die a probability of one-sixth that any particular face is uppermost at the end of a roll (Strevens 1998).

The classical definition of probability, paramount from the time of Gottfried Wilhelm Leibniz to the time of Pierre Simon de Laplace (the late seventeenth century to the early nineteenth century) takes its inspiration from the alignment of probability with symmetry. The best-known formulation of the classical account is due to Laplace:

The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability, which is thus simply a fraction whose numerator is the number of favorable cases and

whose denominator is the number of all the cases possible. (1902, pp. 6–7)

As many commentators note, this formulation, typical of the classical probabilists, appears to involve two parallel definitions, the first based on the notion of equal possibility and the second on the notion of equal undecidedness. Laplace’s relation of equal possibility between two cases probably ought to be understood as picking out a certain physical symmetry in virtue of which the cases have equal physical probabilities. All classical probabilities, on the equal possibility definition, have their basis in such physical symmetries, and so would seem to be physical probabilities. The relation of equal undecidedness between two cases refers to some sort of epistemic symmetry, though perhaps one founded in the physical facts. A probability with its basis in undecidedness would seem to be, by its very nature, an epistemic probability. Classical probability, then, is at the same time a kind of physical probability and a kind of epistemic probability.

This dual nature, historians argue, is intentional (Hacking 1975, Daston 1988). In its epistemic guise classical probability can be called on to do work not normally thought to lie within the province of an objective notion of probability, such as measuring the reliability of testimony, the strength of evidence for a scientific hypothesis, or participating in decision-theoretic arguments such as Blaise Pascal’s famous wager on the existence of God. In its physical guise classical probability is able to cloak itself in the aura of unrevisability and reality that attaches to the gambling probabilities such as the one-half probability of heads.

The classical definition could not last. Gradually, it came to be acknowledged that although the epistemic probabilities may, or at least ought to, shadow the physical probabilities wherever the latter are found, they play a number of roles in which there is no physical probability, nor anything with the same objective status as a physical probability to mimic. The classical definition was split into its two natural parts, and distinct notions of physical and epistemic probability were allowed to find their separate ways in the world.

At first, in the middle and late nineteenth century, physical probability commanded attention almost to the exclusion of its epistemic counterpart. Developments in social science, due to Adolphe Quetelet (1796–1874), in statistical physics, due to James Clerk Maxwell and Ludwig Boltzmann, and eventually (around 1930) in the synthesis of evolutionary biology and genetics, due to Ronald Aylmer Fisher and many others, turned on the successful deployment of physical probability distributions. Begin-

ning in the early twentieth century, however, epistemic probability came into its own, freeing itself over the decades from what came to be seen as the classical probabilists' futile attempt to provide strict guidelines dictating unique rational epistemic probabilities in every situation.

Modern philosophy remade itself in the twentieth century, imposing a historical horizon at around 1900. The story of the interpretation of probability is often told beginning near that year, with the result that the development of epistemic probability, and logical probability in particular, comes first—a convention that will be followed here.

### EPISTEMIC PROBABILITY

Epistemic probability takes two forms. In its first form, it is a measure of a person's degree of confidence in a proposition, increasing from zero to one as his or her attitude goes from almost total disbelief to near certainty. This kind of epistemic probability is called credence, degree of belief, or subjective probability. The propositional attitude one gets when one attaches a subjective probability to a proposition is sometimes called a partial belief.

In its second form, associated most often with the term *logical probability*, epistemic probability measures the impact of a piece or pieces of evidence on a proposition. Its elemental form may not be that of a probability distribution, in the usual sense, but it is related to a probability distribution in some straightforward way, and as will be seen shortly, is quite capable of providing a basis for a complete system of epistemic probability.

There is a foundational dispute between the proponents of the two forms of epistemic probability. It is not a fight for existence but for primacy: The question is which of the two kinds of epistemic probability is the more epistemologically basic.

**LOGICAL PROBABILITY.** The second form of epistemic probability has, since 1900, most often taken the guise of logical probability. A logical probability is attached not to a proposition but to a complete inductive inference. It is a measure of the degree to which the evidence contained in the premises of an inductive inference, considered in isolation, probabilifies the conclusion. The idea of probabilistic inference was an important part of classical probability theory, but from the post-1900 perspective it is associated first with John Maynard Keynes (1921)—who was more famous, of course, as an economist.

In explaining the nature of logical probability, and in particular the tag *logical* itself, Keynes draws a close analogy with deductive inference: Whereas in a deductive inference the premises entail the conclusion, in an inductive inference they partially entail the conclusion, the degree of entailment being represented by a number between zero and one, namely, a logical probability. (Note that a degree zero entailment of a proposition is equivalent to full entailment of the proposition's negation.) Just as the first form of epistemic probability generalizes from belief to partial belief, then, the second form generalizes, in Keynes's hands, from entailment to partial entailment.

For example: Take as a conclusion the proposition that the next observed raven will be black. A proposition stating that a single raven has been observed to be black might entail this conclusion only to a relatively small degree, this logical probability representing the slightness of a single raven's color as evidence for the color of any other raven. A proposition stating that many hundreds of ravens have been observed to be black will entail the conclusion to some much greater degree.

It is an objective matter of fact whether one proposition deductively entails another; so, Keynes conjectured, it is in many cases a matter of objective fact to what degree one proposition partially entails another. These facts themselves comprise inductive logic; the logical probabilities are at base, then, logical entities, just as the name suggests.

Although exact logical probabilities are for Keynes the ideal, he allows that in many cases logic will fix only an approximate degree of entailment for an inductive inference. The presentation in this entry will for simplicity's sake focus on the ideal case.

Keynes's logical probability is not only compatible with subjective probability, the other form of epistemic probability; it also mandates certain values for a person's subjective probabilities. If the premises in an inductive inference are known for certain, and they exhaust the available evidence, then their inductive impact on the conclusion—the degree of entailment, or logical probability attached to the inference, from the premises to the conclusion—is itself the degree of belief, that is, the subjective probability, that a rational person ought to attach to the conclusion, reflecting as it does all and only the evidence for the conclusion.

Keynes uses this argument as a basis for taking as a formal representation of logical probabilities the probability calculus itself: The degree to which proposition *b* entails proposition *a* is written as a conditional probabil-

ity  $P(a|b)$ . Note that these probabilities do not change as the evidence comes in, any more than facts about deductive entailment can change as the evidence comes in. The logical probability  $P(a|b)$  must be interpreted as a quantification of the inductive bearing of  $b$  alone on  $a$ , not of  $b$  together with some body of accepted knowledge.

The unconditional probability  $P(a)$ , then, is the inductive bearing on  $a$  of an empty set of evidence—the degree to which  $a$  is entailed, in Keynes's sense, by the set of logical truths, or tautologies, alone. One might think that the degree of entailment is zero. But this cannot be right: If one has no evidence at all, one must set one's subjective probabilities for both  $a$  and its negation equal to their respective degrees of entailment by the tautologies. But one cannot set both subjective probabilities to zero—it cannot be that one is certain that neither  $a$  nor its negation is true, since one of the two must be true. One's complete lack of evidence would be better represented by setting both subjective probabilities to intermediate values, say one-half. The logical probabilist, in endorsing this assignment, implicitly asserts that the empty set of evidence, or the set of tautologies, entails both  $a$  and its negation to degree one-half.

Although its subject matter is the bearing of evidence on hypotheses, then, logical probability theory finds itself having to take a position on what one should believe when one has no evidence (under the guise of the question of the tautologies' partial entailments). To answer this question, it has turned to the principle of indifference, which recommends—when there is no evidence favoring one of several mutually exclusive possibilities over the others—that the available probability be equally distributed among them. This is, of course, the same principle that comprises one strand of the classical definition of probability: Laplace suggested assigning equal probabilities to cases “such as we may be equally undecided about in regard to their existence” (Laplace 1902, p. 6). It has also played an important role in the development of the theory of subjective probability, and so is discussed in a separate section later in this entry.

As the role of indifference shows, logical probability is close in spirit to the epistemic strand of classical probability. It posits, at least as an ideal, a single system of right reasoning, allowing no inductive latitude whatsoever, to which all rational beings ought to conform. Insofar as rational beings ever disagree on questions of evidential impact, it must be because they differ on the nature of the evidence itself.

Many philosophers find this ideal of inductive logic hard to swallow; even those sympathetic to the idea of

strong objective constraints on inductive reasoning are often skeptical that the constraints take the form of logical truths, or something analogous to logical truths. This skepticism has two sources.

First is the perception that inductive practices vary widely. Whereas there exists a widespread consensus as to which propositions deductively entail which other propositions, there is no such consensus on degrees of evidential support. That is not to say, of course, that there is disagreement about every aspect of inductive reasoning, but there is far less agreement than would be necessary to build, in the same way that deductive logic was constructed, a useful inductive logic.

Second, there are compelling (though not irresistible) reasons to believe that it is impossible to formulate a principle of indifference that is both consistent and strong enough to do the work asked of it by logical probabilists. These reasons are sketched in the discussion of the principle later on.

Rudolf Carnap (1950) attempted to revive the idea of a system of induction founded on logic alone in the mid-century. His innovation—drawing on his general philosophy of logic—was to allow that there are many systems of inductive logic that are, from a purely logical viewpoint, on a par. One may freely choose from these a logic, that is, a set of logical probabilities, that suits one's particular nonlogical ends.

Carnap relativized induction in two ways. First, his version of the principle of indifference was indexed to a choice of language; how one distributes probability among rival possibilities concerning which one knows nothing depends on one's canonical system for representing the possibilities. Second, even when a canonical language is chosen, Carnap's rule for determining inductive support—that this, degrees of entailment or logical probabilities—contains a parameter whose value may be chosen freely. The parameter determines, roughly, how quickly one learns from the evidence. Choose one extreme, and from the observation of a single black raven one will infer with certainty that the next raven will also be black (straight induction). Choose the other extreme, and no number of black ravens is great enough to count as any evidence at all for the blackness of the next raven. A sensible choice would seem to lie somewhere in the middle, but on Carnap's view, logic alone determined no preference ranking whatsoever among the different choices, rating all values apart from the extremes as equally good.



Carnap did give extralogical arguments for preferring a particular value for the parameter, arriving at an inductive rule equivalent to Laplace's rule of succession. Given that, say,  $i$  out of  $n$  observed ravens have been black, both Carnap and Laplace assign a probability of  $(i + 1)/(n + 2)$  to the proposition that the next raven will be black.

One awkward feature of Carnap's system is that, no matter what value is chosen for the inductive parameter, universal generalizations cannot be learned: The inductive bearing of any number of black ravens on the hypothesis "All ravens are black" is zero.

Carnap's system is of great intrinsic interest, but from the time of its presentation, its principal constituency—philosophers of science—was beginning to move in an entirely different direction. Such considerations as Nelson Goodman's new riddle of induction and arguments by Bayesians and others that background knowledge played a part in determining degrees of inductive support, though not beyond the reach of Carnap's approach, strongly suggested that the nature of inductive support could not be purely logical.

Today, the logical approach to inductive inference has been supplanted to a great extent by (though not only by) the Bayesian approach. Still, in Bayesianism itself some have seen the seeds of a new inductive logic.

**SUBJECTIVE PROBABILITY.** Whereas logical probability is a logical entity—a quantification of the supposed logical facts about partial entailment—the other kind of epistemic probability, subjective probability, is a psychological entity, reflecting an actual cognitive fact about a particular person or (if they are sufficiently agreed) a group of people. The rationality of a person's subjective probabilities may be a matter of logic, then, but the probabilities themselves are a matter of psychology.

That for a number of propositions one tends to have a degree of confidence intermediate between the extremes associated with total disbelief and total belief, no one will deny. The advocates of subjective probability as a key epistemological notion—who call themselves Bayesians or simply subjectivists—go much further than this. They characteristically hold that humans have, or ought to have, well-defined subjective probabilities for every proposition and that these subjective probabilities play a central role in epistemology, both in inductive inference, by way of Thomas Bayes's (1702–1761) conditionalization rule, and in practical deliberation, by way of the usual mechanisms of decision theory.

The subjectivist's first challenge is to give a substantial characterization of subjective probability and to argue that subjective probabilities are instrumental in human cognition, while at the same time finding a foothold in the descriptive, psychological scheme for the normative concerns of epistemology. Much of this groundwork was laid in Frank Plumpton Ramsey's influential paper "Truth and Probability" (1931).

Ramsey does not define subjective probability as such, and even goes so far as to acknowledge that the ideal of a definite subjective probability for every proposition is just that—an ideal that goes a long way toward capturing actual human epistemology without being accurate in every respect. What he posits instead is a connection—whether conceptual or empirical he does not say—between the value of a person's subjective probability for a proposition and his or her betting behavior.

If one has a subjective probability  $p$  for a proposition  $a$ , Ramsey claims, one will be prepared to accept odds of up to  $p$ :  $(1 - p)$  on the truth of  $a$ . That is, given a game in which one stands to win  $\$n$  if  $a$  is true, one will pay up to  $\$pn$  to play the game; equivalently, if one will pay up to  $\$m$  to play a game in which one stands to win  $\$n$  if  $a$  is true, one's subjective probability for  $a$  must be  $m/n$ . (Decision theorists, note, talk about utility, not dollars.)

Importantly, all human choice under uncertainty is interpreted as a kind of betting. For example, suppose I have to decide whether to wear a seat belt on a long drive. I am in effect betting on whether I will be involved in an auto accident along the way. If the cost of wearing a belt, in discomfort, inconvenience, and forsaken cool, is equivalent to losing  $\$m$ , and the cost of being beltless in an accident, in pain, suffering, and higher insurance premiums, is  $\$n$ , then I will accept the risk of going beltless just in case my subjective probability for there being an accident is less than or equal to  $m/n$ . (Here, the "prize" is negative. The cost of playing is also negative, so just by agreeing to play the game, I gain something: the increase in comfort, cool, and so on. My aim is to play while avoiding a win.) The central doctrine of decision theory is, then, built into the characterization of subjective probability.

Ramsey (1931) uses this fact to argue that, provided a person's behavior is coherent enough to be described, at least approximately, by the machinery of decision theory, his or her subjective probabilities for any proposition may be inferred from his or her choices. In effect, the person's subjective probabilities are inferred from the nature of the bets, in the broadest sense, he or she is prepared to accept. Because one's overt behavior can be systematized,

approximately, using a decision-theoretic framework, one must have subjective probabilities for every proposition, and these probabilities must play a central role in one's decision theory.

What is the force of the *must* in the preceding sentence? That depends on the nature of the posit that one having a certain subjective probability for a proposition means that one is prepared to accept certain odds on the proposition's being true. Some writers, especially in the midcentury heyday of conceptual analysis and psychological behaviorism, interpret the posit as a definition of subjective probability; on this view, one having certain subjective probabilities just is one having a certain betting behavior. Others, like Ramsey (1931), opt for a looser connection. On any approach, there is a certain amount of latitude in the phrase "prepared to accept." If I am prepared to accept certain odds, *must* I play a game in which I am offered those odds? Or only if I am in a betting mood? The former answer vastly simplifies the subjectivist enterprise, but at a cost in psychological plausibility: It is surely true that people frequently gamble in the broad sense that they take measured risks, but it is not nearly so obvious that they are compulsive gamblers intent on taking on every favorable risk they can find. Work on the psychology of decision making also suggests that it is a mistake to found the subjectivist enterprise on too strong a conception of the connection between subjective probability and betting behavior.

Subjective probabilities are supposed to conform, as the name suggests, to the axioms of probability theory. In a theory such as Ramsey's (1931), a certain amount of probability mathematics is built into the technique for extracting the subjective probabilities; that humans not only have subjective probabilities, but arrange them in accord with the axioms, is a condition for the success of Ramsey's (1931) project.

Insofar as subjective probability is not simply defined as whatever comes out of the Ramsey project, however, there is a question whether subjective probabilities obey the axioms. If they do not, there is little that they are good for, so the question is an important one for subjectivists, who tend to follow Ramsey in giving a normative rather than a descriptive answer: It is rational to arrange one's subjective probabilities in accordance with the axioms. (It is not unreasonable, of course, to see this normative claim, if true, as evidence for the corresponding descriptive claim, since humans are in certain respects reliably rational.)

The vehicle of Ramsey's argument is what is called the Dutch book theorem: It can be shown that, if one's

subjective probabilities violate the axioms, then one will be prepared to accept certain sets of bets (which bets depends on the nature of the violation) that will cause one a sure loss, in the sense that one will lose whether the propositions that are the subjects of the bets turn out to be true or false.

The details of the argument are beyond the scope of this entry (for a more advanced introduction, see Howson and Urbach 1993), but an example will illustrate the strategy. The axioms of the probability calculus require that the probability of a proposition and that of its negation sum to one. Suppose one violates this axiom by assigning a probability of 0.8 both to a certain proposition  $a$  and to its negation. Then one is prepared to accept odds of 4:1 on both  $a$  and  $\neg a$ , which means a commitment to playing, at the same time, two games, in one of which one pays \$8 and wins \$10 (i.e., one's original \$8 plus a \$2 profit) if  $a$  is true, and in one of which one pays \$8 and wins \$10 if  $a$  is false. Whether  $a$  is true or false, one pays \$16 but wins only \$10—a certain loss. To play such a game is irrational; thus, one should conform one's subjective probabilities to the probability calculus. Needless to say, the Dutch book argument works best on the dubious interpretation of "prepared to accept" as equivalent to "compelled to accept"; there have been many attempts to reform or replace the argument with something that makes weaker, or even no, assumptions about betting behavior.

Subjectivism has been developed in several important directions. First are various weakenings or generalizations of the subjectivist machinery. The question of the connection between subjective probability and betting behavior is, as noted, one locus of activity. Another attempts to generalize the notion of a subjective probability to a subjective probability interval, the idea being that where one does not have an exact subjective probability for a proposition, one may have an approximate level of confidence that can be captured by a mathematical interval, the equivalent of saying that one's subjective probability is indeterminately somewhere between two determinate values.

Second, and closely related, is all the work that has been put into developing decision theory over the last 100 years (e.g., see Jeffrey 1983). Finally, subjectivism provides the foundation for the Bayesian theory of inference. At the root of the Bayesian system is a thought much like the logical probabilist's doctrine that, if  $k$  is one's background knowledge, then one's subjective probability for a hypothesis  $a$  ought to be  $P(a|k)$ . Whereas for a logical probabilist a conditional probability  $P(a|b)$  is a timeless

logical constant, for a subjectivist it is something that constantly changes as further evidence comes in (even holding  $a$  and  $b$  fixed). For this reason, the subjectivist theory of inference must be an inherently dynamic theory; what is perhaps its best-known weakness, the “problem of old evidence,” arises from this fact.

Subjectivism had almost entirely eclipsed logical probabilism by the late twentieth century; as the celestial metaphor unwittingly implies, however, there is a cyclic aspect to philosophical history: An interest in the central notion of logical probability theory, evidential weight, is on the rise.

There are three strands to this new movement. First is the perception among philosophers of science that scientific discourse about evidence is almost never about the subjective probability scientists should have for a hypothesis, and almost always about the degree of support that the evidence lends to the hypothesis. Second is the development of new and safer (though limited) versions of the principle of indifference. Third is technical progress on the project of extracting from the principles of Bayesian inductive inference a measure of weight. Note that this third project conceives of inductive weight as something derived from the more basic Bayesian principles governing the dynamics of subjective probability, a view opposed to the logical probabilists’ derivation of rational subjective probabilities from the (by their lights) more basic logical principles governing the nature of inductive support.

**INDIFFERENCE.** The principle of indifference distributes probability among various alternatives—in the usual case, mutually exclusive and exhaustive propositions—concerning which little or nothing is known. The principle’s rationale is that certain probability distributions reflect ignorance better than others. If I know nothing that distinguishes two mutually exclusive possibilities, picked out by propositions  $a$  and  $b$ , then I have no reason to expect one more than the other: I should assign the propositions equal probabilities. Any asymmetric assignment, say assigning twice the probability to  $a$  that I assign to  $b$ , would reflect some access on my part to facts supporting  $a$  at the expense of  $b$ . Thus, ignorance and probabilistic symmetry ought to go hand in hand—or so the principle of indifference would have it.

The principle is an essential part of logical probability theory, for the reasons given earlier, but there have always been subjectivists who appeal to the principle as well. It is most useful within the Bayesian approach to inductive inference.

The epistemic strand of classical probability theory also invokes the principle, of course, blending it with the discernment of “equally possible cases” in the paradigmatic gambling setups. This conflation has confused the discussion of the principle ever since, with proponents of the principle continuing to take aid and comfort in the principle’s apparent virtuoso handling of cases such as the one-half probability of heads. One’s reasoning about the gambling probabilities, however, as the classical probabilists for the most part themselves dimly saw, is a matter of inferring physical probabilities from physical symmetries, not of setting epistemic probabilities to reflect symmetric degrees of ignorance (Strevens 1998).

The most famous arguments against the principle of indifference were developed in the nineteenth century, which was a time of hegemony for physical over epistemic probability. They take their name from Joseph Bertrand (1822–1900), who pointed to the difficulty of finding a unique symmetry in certain indifference-style problems.

Consider, for example, two leading theories of dark matter in the universe: the MACHO and the WIMP theories. Each posits a certain generic form for dark matter objects, respectively large and small. If one has no evidence to distinguish them, it seems that the principle of indifference directs one to assign each a probability of one-half (assuming for the sake of the argument that there are no other possibilities). But suppose that there are four distinct schools of thought among the MACHO theorists, corresponding to four distinct ways that MACHOs might be physically realized, and eight such schools of thought among WIMP theorists. Now there are twelve possibilities, and once probability is distributed equally among them, the generic MACHO theory will have a probability of one-third and the WIMP theory a probability of two-thirds. Cases such as this make the principle seem capricious, if not simply inconsistent (as it would be if it failed to pick out a privileged symmetry).

Matters become far worse, as Bertrand noted, when there are uncountably many alternatives to choose among, as is the case in science when the value of a physical parameter, such as the cosmological constant, is unknown. Even in the simplest of such cases, the principle equivocates (Van Fraassen 1989, chapter 12). As noted earlier, some progress has been made in solving these problems, with Edwin T. Jaynes (1983) being a ringleader. Most philosophers, though, doubt that there will ever be a workable principle of indifference suited to the needs of general inductive inference.

## PHYSICAL PROBABILITY

The paradigms of physical probability are the probabilities attached to gambling setups; there are, however, many more interesting examples: the probabilities of quantum mechanics and kinetic theory in physics, the probabilities of population genetics in evolutionary theory, actuarial probabilities such as the chance of dying before reaching a certain age, and the probabilities in many social science models. It is by no means clear that there is a single phenomenon to be explained here; the physical probabilities ascribed to phenomena by the best scientific theories may differ in their makeup from theory to theory. There is a commonality in the phenomena themselves, however: Whenever the notion of physical probability is put to scientific work, it is to predict or explain what might be called probabilistic patterns of outcomes. These patterns are characterized by a certain kind of long-run order, discernible only over a number of different outcomes, and a certain kind of short-term disorder, the details of the order and disorder depending on the variety of probability distribution.

The simplest and best-known of the patterns is the Bernoulli pattern, which takes its name from the corresponding probability distribution. This is the pattern typical of the outcomes produced by gambling devices, such as the pattern of heads and tails obtained by tossing a coin. The long-term order takes the form of a stable frequency equal to the corresponding probability. In the case of the tossed coin, this is of course the one-half frequency with which heads and tails occur (almost always) in the long run. The short-term disorder, though an objective property of the pattern itself, is perhaps best gotten at epistemically: Once one knows that the long-run frequency of heads is one-half, the outcome of one toss provides no useful information about the outcome of the next. The law of large numbers implies that a chance setup will produce its characteristic probabilistic patterns in the long run with very a high (physical) probability. When discussing physical probability, it is more natural to talk of probabilities attaching to events than to propositions; what follows will be formulated accordingly.

**THE FREQUENCY THEORY** The frequentist theory of physical probability has its roots in the empiricist interpretation of law statements according to which they assert only the existence of certain regularities in nature (on the regularity theory, see Armstrong 1983). What is usually called the actual frequency theory of probability understands physical probability statements, such as the claim that the probability of a coin toss's yielding heads is

one-half, as asserting in a like spirit the existence of the appropriate probabilistic patterns—in the case of the coin toss, for example, a pattern of heads and tails in the actual outcomes of coin tosses exemplifying both the order and the disorder characteristic of the Bernoulli patterns.

The characteristic order in a Bernoulli pattern is a long-run frequency approximately equal to the relevant probability; in the case of the coin, then, it is a long-run frequency for heads of one-half. It is from this aspect of the pattern that frequentism takes its name. (One complication: A distinction must be made between the case in which the set of events exemplifying the pattern is finite and the case in which it is countably infinite. In a finite case, what matters is the proportion or relative frequency, whereas in the infinite case, it is instead the limiting frequency, that is, the value of the relative frequency in the limit, if it exists, as it must for the Bernoulli pattern to exist.)

Although their account is named for frequencies, most frequentists insist also on the presence of appropriate short-term disorder in the patterns. It is less easy to characterize this disorder in the purely extensional terms implicit in a commitment to regularity metaphysics. Suffice it to say that there is a broad range of characterizations, some strict, some rather lax. Among frequentists, Richard von Mises (1957) tends to a strict and Hans Reichenbach (1949) to a lax requirement (though Reichenbach holds, characteristically, that there is no uniquely correct level of strictness; for a discussion of the technical problems in constructing such a requirement, see Fine [1973]).

The probability that a particular coin toss lands heads is one-half, according to frequentism, because the outcome of the toss belongs to a series that exemplifies the Bernoulli pattern with a frequency of one-half. The truth-maker for the probability claim is a fact, then, about a class of outcomes, not just about the particular outcome to which the probability is nominally attached. But which class? If one is tossing an American quarter, does the class include all American quarters? All American and Canadian quarters? All fair coins? Or—ominously—all coin tosses producing heads? To give an answer to this question is to solve what has become known as the problem of the reference class.

The standard frequentist solution to the problem is to understand probability claims as including a (perhaps implicit) specification of the class. All physical probability claims are, in other words, made relative to a reference class. This doctrine reveals that the frequency theory is

best seen as an account, in the first instance, of statements of statistical laws. A claim about the one-half probability of heads, for example, is on the frequency interpretation in essence a statement of a probabilistic law concerning a class of coin tosses, not a claim about a property of a particular toss.

The kinship between the regularity account of deterministic laws and the frequency account of probability is, then, even closer than it first appears. Note that the regularity account has its own analog of singular probability claims, namely, singular claims about deterministic tendencies, such as a particular brick's tendency to fall to earth when released. Regularity theorists interpret a tendency claim not as picking out an intrinsic property of the object possessing the tendency, but as a veiled law statement.

The case of probability introduces a complication, however, that is not present in the case of exceptionless regularities: A particular coin toss will belong to many reference classes, some with different frequencies for heads. There may be, then, no determinate fact of the matter about an individual coin toss's probabilistic tendency to produce heads, or equivalently, about what are often called single case probabilities. Frequentists have made their peace with this consequence of their view.

Opponents of the frequency view argue that single-case probabilities are metaphysically, inductively, and explanatorily indispensable. Are they right? Here is the case for metaphysical indispensability: Some writers, especially propensity theorists, hold that there is clearly a fact of the matter about the value of the probability that some particular coin toss lands heads, independent of any choice of reference class. Frequentists may simply deny the intuition or may try explain away the appearance of a single-case fact (for related versions of the explanation, see Reichenbach 1949, §68; Strevens 2003, pp. 61–62).

And here is the case for predictive indispensability: To settle, for predictive and decision-theoretic purposes, on a rational subjective probability for an event using the probability coordination principle, a corresponding physical probability must be found (see the discussion of probability coordination later on). The corresponding probability is often understood to be the physical probability of that very event, hence, a single-case probability. Frequentists must find an alternative understanding. Reichenbach proposes using the frequentist probability relative to the narrowest reference class “for which reliable statistics can be compiled” (1949, p. 374).

The case for explanatory indispensability rests principally on the intuition that the probabilistic explanation of a single outcome requires a single-case probability. The philosophy of scientific explanation, much of it developed by regularity theorists and other metaphysical empiricists, offers a number of alternative ways of thinking about explanation, for example, as a matter of showing that the outcome to be explained was to be expected, or as a matter of subsuming the outcome to be explained under a general pattern of outcomes (both ideas proposed by Carl Gustav Hempel). The fate of frequentism, and more generally of the regularity approach to laws of nature, depends to some extent, then, on the adequacy of these conceptions of explanation.

Why be a frequentist? The view has two principal advantages. First is its light metaphysical touch, shared with the regularity account of laws. Second is the basis it gives for the mathematics of probability: Frequencies, as mathematical objects, conform to almost all the axioms of probability. Only almost all because they violate the axiom of countable additivity, an extension to the countably infinite case of the third axiom described earlier. Countable additivity plays an important role in the derivation of some of probability mathematics' more striking results, but whether it is necessary to provide a foundation for the scientific role of physical probability claims is unclear.

There is more than one way to be a frequentist. A naive actual frequentist holds that there is a probability wherever there is a frequency, so that, in a universe where only three coin tosses have ever occurred, two coming up heads, there is a probability for heads of two-thirds. This view has been widely criticized, though never held. Compare with the naive regularity theory of laws (Armstrong 1983, §2.1).

What might be called ideal actual frequentism is the theory developed by Reichenbach (1949) and von Mises (1957). On this view, probability statements are construed as ideally concerning only infinite classes of events. In practice, however, they may be applied to large finite classes that in some sense come close to having the properties of infinite classes. Thus, Reichenbach distinguishes the logical meaning of a probability statement, which asserts the probabilistic patterning of an infinite class of outcomes, and the finitist meaning that is given to probability claims in physical applications, that is, in the scientific attribution of a physical probability (Reichenbach 1949). On the finitist interpretation, then, a physical probability claim concerns the probabilistic patterning of some actual, finite class of events—albeit a class large

enough to have what Reichenbach calls a practical limiting frequency. (Reichenbach's wariness about logical meaning owes as much, incidentally, to his desire to have his theory of probability conform to the verifiability theory of meaning as to a concern with, say, the validity of probability claims in a finite universe.)

David Lewis (1994), reviving Ramsey's account of laws of nature, proposes that the fundamental laws are nothing but the axioms of the theory that best systematizes, or unifies, the phenomena. A systematization is good to the degree that it is simple, that it makes claims about a large proportion of the phenomena (ideally all the phenomena, of course), and that its claims are accurate. Lewis (1994) extends the definition of accuracy, or as he calls it, fit, to accommodate axioms attributing physical probabilities: A set of phenomena are a good fit to a physical probability statement if the phenomena exemplify the probabilistic patterns appropriate to the probability ascribed. A system of probabilistic axioms will be a good systematization, then, only if the physical probabilities it assigns to the phenomena are reflected, for the most part, in corresponding probabilistic patterns.

In this respect, Lewis's view is a form of frequentism. Although there is not some particular set of outcomes whose probabilistic patterning is necessary and sufficient for the truth of a given probabilistic law statement, it is nevertheless the world's probabilistic patterns, taken as a whole, that provide the basis for all true statements of probabilistic law.

Some writers suggest that a claim such as "The probability of obtaining heads on a toss of this coin is one-half" is equivalent to the claim that, if the coin were tossed infinitely many times, it would yield heads with a limiting frequency of one-half. The truth-makers for physical probability claims, then, are modal facts (except in the case where there actually are an infinite number of tosses). This view is known as hypothetical frequentism.

Though much discussed in the literature, hypothetical frequentism is seldom advocated. Reichenbach (1949) and von Mises (1957) are sometimes labeled hypothetical frequentists, but the textual evidence is thin, perhaps even nonexistent. Colin Howson and Peter Urbach (1993) advocate a hypothetical frequency view. Bas C. van Fraassen's (1980) frequencies are also hypothetical, but because he holds that the literal meaning of theoretical claims is irrelevant to the scientific enterprise, the spirit of his account of probability is, in its empiricism, closer to Reichenbach's ideal actual frequentism.

The weaknesses of frequentism are in large part the weaknesses of the regularity theory of laws. An interesting objection with no parallel in the regularity account is as follows: In the case of reference classes containing countably infinite numbers of events, the value (indeed, the existence) of the limiting frequency will vary depending on how the outcomes are ordered. There appear to be no objective facts, then, about limiting frequencies. Or rather, if there are to be objective facts, there must be some canonical ordering of outcomes, either specified along with the reference class or fixed as a part of the scientific background. How serious an impediment this is to the frequentist is unclear.

**THE PROPENSITY THEORY.** If frequentism is the regularity theorist's natural interpretation of physical probability claims, then the propensity account is the interpretation for realists about laws, that is, for philosophers who believe that law statements assert the existence of relations of nomic necessity and causal tendencies (Armstrong 1983). For the propensity theorist, probabilities are propensities, and propensities are a certain kind of distinctly probabilistic causal tendency or disposition.

The propensity theorist's home territory is single-case probability, the kind of probability attached to a particular physical process or outcome independently of the specification of a reference class or ordering of outcomes. Because propensities are supposed to be intrinsic properties of token processes, on the propensity view every probability is a single-case probability. Given some particular outcome that one wishes to predict or explain, then, there is an absolute fact of the matter as to the physical probability of the outcome that one may—and presumably, must—use in one's prediction or explanation.

Of course, knowledge of this fact, if it is to be obtained by observing the statistics of repeated experiments, will require the choice of a reference class, the aim being to find a class containing processes that are sufficiently similar that their statistics reveal the nature of each of the underlying propensities in the class. Furthermore, by analogy with the case of deterministic causal tendencies, propensities may owe their existence to probabilistic laws governing classes of processes. Thus, something not unlike the frequentist's reference classes may turn up in both the epistemology and the metaphysics of propensities, but this does not detract from the fact that on the propensity view, there are real, observer-independent single-case probabilities.

To identify probabilities with propensities is revealing because one thinks that one has a good intuitive sense

of the nature of propensities in the deterministic case; one is reasonably clear on what it is to be fragile, aggressive, or paramagnetic. Though the metaphysics of dispositions is still a matter of dispute, it seems that one comes to deterministic propensities, at least at first, by grasping what they are propensities for: for example, breaking, violent behavior, and magnetic attraction. To adopt a propensity theory of probability, then, with the sense of familiarity the word *propensity* brings, is to make an implicit commitment to elucidating what probabilistic propensities are propensities for.

A straightforward answer to this question was given by Karl R. Popper (1959) in one of the earliest modern presentations of the propensity theory: A probabilistic propensity is a disposition to produce probabilistically patterned outcomes. A particular coin's probability for heads of one-half, then, is a disposition to produce a sequence of heads and tails that is disordered in the short term, but in the long term contains heads with a frequency of one-half. (Popper in fact omits the disorder requirement and allows that the sequence may be long and finite or infinite.) On Popper's view, then, a probabilistic propensity differs from a deterministic propensity not in the means of production, but only in what is produced: a probabilistic pattern over a long series of trials, rather than a single discrete episode of, say, shattering or magnetic attraction.

Popperian propensity theory is committed to the claim that, if the probability of a tossed coin's landing heads is one-half (and remains so), then continued tossing of the coin will eventually yield a set of outcomes of which about one-half are heads. But this sits badly with the intuitive conception of the workings of probability: If the probability of heads is one-half, then it is possible, though unlikely, that it will produce all heads for as long as one likes, even forever.

This intuition has an analog in probability mathematics. The law of large numbers prescribes a very high probability that the long-run frequency with which an outcome occurs will match its probability; by the same token, however, there is a nonzero probability that any (finite) long run will fail to produce a probability-matching frequency. There is some physical probability, then, that a probabilistic propensity will fail to produce what, according to the Popperian propensity view, it must produce. If this physical probability is itself a Popperian propensity—and surely it is just another manifestation of the original one-half propensity for heads—then it must produce, by Popper's definition, a matching frequency, which is to say that it must

occasionally produce the supposedly impossible series of heads. If it is to be consistent, Popper's definition must be carefully circumscribed. (There is a lesson here for frequentists, too.)

Most propensity theorists accept that probabilistic setups will occasionally fail to produce probability-matching frequencies. Thus, they repudiate Popper's version of the propensity theory. What, then, can they say about the nature of the propensity? Typically, they hold that the probability of, say, heads is a propensity to produce the appropriate probabilistic patterns with a high physical probability (Fetzer 1971, Giere 1973)—thus, such a probabilistic propensity is probabilistic not only in its characteristic effect, which is, as on Popper's definition, a probabilistic pattern, but also in its relation to the effect. (D. H. Mellor [1971] offers an interesting variant on this view.)

Whereas the Popperian definition comes close to inconsistency, this new definition is manifestly circular. Its proponents accept the circularity, so committing themselves to the ineffability of probabilistic propensities.

The ineffability of propensities, it is asserted, is not a problem provided that their values can be inferred; the usual apparatus of statistical inference is tendered for this purpose. Critics of the post-Popperian propensity interpretation naturally fasten on the question of whether it succeeds in saying anything substantive about probability at all—anything, for example, that illuminates the question of why physical probabilities conform to the axioms of the probability calculus or explain the outcomes that they produce. It does seem that modern propensity theorists are not so far from what is sometimes called the semantic interpretation of probability, on which probabilities are considered to be model-theoretic constructs that ought not to be interpreted at all, but simply accepted as formal waypoints between evidence and prediction in probabilistic reasoning (Braithwaite 1953). Compare Carnap's (1950) notion of partial interpretation and Patrick Suppes (1973).

A characteristic doctrine of the propensity theory is that probabilistic propensities, hence probabilities, are metaphysically irreducible: They are in some sense fundamental building blocks of the universe. The corollary to this doctrine is that the physical probabilities science assigns to outcomes that are deterministically produced—including, according to many philosophers, the probabilities of statistical mechanics, evolutionary biology, and so on—are, because they are not irreducible, they are not propensities, and because they are not

propensities, they are irreducible. Ronald N. Giere (1973) writes that they must be given an “as if” interpretation, but propensity theorists offer no account of “as if” probability’s scientific role.

On a broader understanding of the nature of a propensity, however, at least some of the physical probabilities assigned by science to the outcomes of deterministic processes might count as probabilistic propensities. As explained in the entry on chaos, certain subclasses of chaotic systems have dynamic properties in virtue of which they tend to generate probabilistic patterns of outcomes (Strevens 2003). These dynamic properties may be understood, then, as endowing the systems with a propensity to produce probabilistic patterns, and the propensity itself may be identified with the physical probabilities that science ascribes to the outcomes.

There is one, not inconsiderable, complication: The systems in question will generate the probabilistic patterns only given appropriate initial conditions. Almost all, but not all, initial conditions will do. This raises two important questions that need to be answered if chaos is to provide a part of the foundation for the metaphysics of physical probability. First, ought the necessary properties of the initial conditions to be considered a part of the propensity? If so, the propensity seems not to be an intrinsic causal property of the process. Second, the initial conditions are, in this context, most naturally described using a probability distribution. Thus, the basis of the probabilistic propensity is a further probabilistic element itself in need of analysis.

**THE SUBJECTIVIST THEORY.** It is something of a mystery why the mathematics of the probability calculus should be useful both for capturing elements of belief and inductive inference and for describing the processes that give rise to probabilistic patterns, or in other words, why two such different things as epistemic and physical probability should share the same formal structure.

According to the subjectivist theory of physical probability, there is no mystery at all: Physical probabilities are nothing but a certain kind of subjective probability. The intuition that, say, the probability of heads is a quantification of some physical property of the tossed coin is, on the subjectivist approach, an illusion: There are frequencies and mechanical properties out in the world, but physical probabilities exist entirely in the descriptive apparatus of people’s theories, or in their minds.

For the principal architect of subjectivism, Bruno de Finetti, the appeal of the theory is not only its neoclassical reunification of epistemic and physical probability but

also its empiricism: Subjectivism is most at home in what is now called a Humean world. Of course, frequentism is also a theory of physical probability that the metaphysical empiricist can embrace; the main advantage of subjectivism over frequentism is its provision—if such is truly necessary—of single-case probabilities (de Finetti 1964).

Subjectivism asserts the identity of the subjective probability for heads and the physical probability for heads. But it does not claim that, say, one’s subjective probability for the MACHO theory of dark matter is also a physical probability for the theory. Rightly so, because one does not acknowledge the existence of physical probabilities wherever there are subjective probabilities. A plausible subjectivism must have the consequence that one projects only a small subset of one’s subjective probabilities onto the world as physical probabilities.

At the heart of the subjectivist theory, then, must be a criterion that picks out just those subjective probabilities that are experienced as physical and that accounts for their particular, peculiar phenomenology. The key notion in the criterion is one of resilience: Unlike most subjective probabilities, which change as more evidence comes in, the subjective probabilities one calls physical have attained a certain kind of stability under the impact of additional information. This stability gives them the appearance of objectivity, hence of reality, hence of physicality, or so the subjectivist story goes. Brian Skyrms (1980) employs this same notion of resilience to give a projectivist account of causal tendencies and lawhood in the deterministic as well as the probabilistic case; subjectivism, then, like frequentism and the propensity theory, can be seen as a part of a larger project embracing all causal and nomological metaphysics.

There is an obvious difficulty with the subjectivist position as elaborated so far: My subjective probability for an outcome such as a coin’s landing heads may very well change as the evidence comes in. I may begin by believing that a certain coin is fair, and so that the physical probability of its yielding heads when tossed is one-half. As I continue to toss it, however, I may come to the realization that it is biased, settling eventually on the hypothesis that the physical probability of heads is three-quarters. Throughout the process of experimentation, I project (according to the subjectivist) a physical probability distribution onto the coin, yet throughout the process, because the projected physical probability for heads is changing, increasing from one-half to three-quarters, my subjective probability for heads is also changing. Where is the resilience?



De Finetti's (1964) achievement is to find a kind of resilience, or constancy, in my subjective probabilities even as my subjective probability for heads is changing. This resilience is captured by the property de Finetti calls exchangeability. Consider my subjective probability distribution over, say, the outcomes of the next four tosses of my coin. Every possible sequence of four outcomes will be assigned some subjective probability. The probability assignment—the subjective probability distribution—is said to be exchangeable if any two sequences having the same number of heads and tails are assigned equal probabilities. For example, exchangeability implies that HTHT and HHTT, each having two heads and two tails, are assigned the same probability, but allows this probability to differ from that assigned to, say, HHHT. In an exchangeable distribution, then, the probability assigned to a sequence of heads and tails depends only on the relative frequency with which heads and tails occur in the sequence (in the case of infinite sequences, which de Finetti uses in his mathematical construction, substitute *limiting frequency*).

If my subjective probability distribution over heads and tails is exchangeable, then the order in which the heads and tails come in as I experiment with my coin will not in itself affect my subjective probability for heads. The frequency with which heads and tails come in will, by contrast, most definitely affect my subjective probability. Thus, exchangeability is a kind of partial resilience; it is resilience to information about order, but not frequency.

De Finetti (1964) claims, uncontroversially, that one's subjective probability distributions over future sequences of heads and tails (and the outcomes of other Bernoulli setups) are exchangeable. He goes on to prove a theorem—his celebrated representation theorem—that shows that the following two reasoners will be outwardly indistinguishable: First, a reasoner who has various hypotheses about the physical probability of heads and updates the subjective probabilities for these hypotheses in the usual way as evidence comes in, and second, a reasoner who has no beliefs about physical probabilities, but simply has an exchangeable subjective probability distribution over future sequences of outcomes. The only difference between the two reasoners, then, will be that the first will claim, presumably as a result of introspection, to be learning about the values of physical probabilities in the world.

The subjectivist's sly suggestion is that people are all in fact reasoners of the second kind, falsely believing that they are reasoners of the first kind. Or, in a more revisionist mood the subjectivist may argue that, though they are reasoners of the first kind, they will give up nothing

but dubious metaphysical commitments by becoming reasoners of the second kind.

Critics of subjectivism question the aptness of exchangeability as a psychological foundation for probabilistic reasoning. The sole reason that people assign exchangeable subjective probability distributions to certain classes of sequences, according to these writers, is that they believe the sequences to be produced by physical probabilities (Bernoulli distributions, to be exact) and they know that an exchangeable subjective probability distribution is appropriate for outcomes so produced. Note that this argument has both a descriptive and normative dimension: Against a descriptive subjectivist, who holds that beliefs about physical probability play no role in people's probabilistic reasoning, the critic proposes that such beliefs cause them to assign exchangeable distributions. Against a normative subjectivist, who holds that beliefs about physical probability should not play a role in people's probabilistic reasoning, the critic proposes that such beliefs are required to justify their assigning exchangeable distributions.

A different line of criticism targets subjectivism's metaphysics: Why not identify physical probability with whatever produces the probabilistic patterns? Why not say that the probability of heads is a quantification of, at least in part, the physical symmetry of the coin? Such a position has its problems, of course, but they are not obviously insurmountable. More generally, given the rich array of options available for understanding the nature of physical probability, the subjectivist's flight from any attempt to give a metaphysics seems to many, as yet, insufficiently motivated.

## PROBABILITY COORDINATION

It is generally accepted that it is rational, in normal circumstances, to set one's subjective probability for an event equal to the physical probability ascribed by science to that event or to that type of event. Returning to the first paragraph of this entry, if the physical probability of a hurricane is high, I should expect—I should assign a high subjective probability to—a hurricane strike. This is the principle of probability coordination.

Because the equation of physical and epistemic probability is made explicit in the classical definition of probability, classicists are probability coordinators par excellence. Leibniz, for example, articulates what appears to be an early formulation of the probability coordination principle when he writes “quod facile est in re, id probabile est in mente” (Hacking, 1975, p. 128); Ian Hacking glosses this as “our judgment of probability ‘in the mind’

is proportional to (what we believe to be) the facility or propensity of things” (the parenthesized phrase is not in the Latin; 1975, p. 128). But strictly speaking, of course, classicists cannot conceive of this as a coordination of different kinds of probability, since they allow only one kind of probability.

In the twentieth century, probability coordination was introduced as a topic in its own right by David Miller, who argued, as a part of a Popperian case against inductive inference, that a probability coordination principle would have to be inconsistent. Commentators soon pointed out that there are consistent versions of the principle, and some years later David Lewis wrote what is still the most influential paper about the proper form of a principle of coordination and its role in scientific inference, conjecturing that such a principle “capture[s] all we know about [physical probability]” (1980, p.266).

Modern attempts at a formulation of a probability coordination principle contain two elements not present in Leibniz’s maxim. First is the modification interpolated by Hacking: The principle commands that one sets one’s subjective probabilities equal not to the corresponding physical probabilities, but to what one believes the values of those probabilities to be, or more generally, to the mean of the different possible values, weighted by one’s subjective probability that each value is the correct one. Such a principle might be loosely interpreted as saying that one should do one’s best to set one’s subjective probabilities equal to the physical probabilities.

Second is a restriction of the range of the principle: When one possesses certain kinds of information, probability coordination is not necessarily rational. Suppose, for example, that I know for some science-fictional reason that the coin I am about to toss will land heads. Then I should set my subjective probability for heads equal to one, not equal to the physical probability of one-half. The information that the coin will land heads is what Lewis (1980) calls inadmissible information; in the presence of inadmissible information, the principle of probability coordination does not apply. Note that what is admissible is relative to the outcome in question; knowing how the coin lands is admissible when I am setting my subjective probability for the outcome of a different toss.

An attempt at a probability coordination principle might, then, have the following form: one’s subjective probability for an event  $e$ , conditional both on the proposition that the physical probability of  $e$  is  $p$  and on any admissible information  $k$ , should be set equal to  $p$ . (One’s unconditional subjective probability for  $e$ , then, will be the weighted sum of the physical probabilities, as men-

tioned earlier.) In symbols: If one’s background knowledge is admissible, then set

$$C(e|tk) = P_t(e),$$

where  $C(\cdot)$  is one’s subjective probability distribution,  $t$  is the proposition that the correct physical probability distribution for  $e$  is  $P_t(\cdot)$ , and  $k$  is any other admissible information.

Note that propositions such as  $t$  are normally consequences of two kinds of fact: probabilistic laws of nature and some properties of  $e$  in virtue of which it falls under the laws. For example, if  $e$  is the event of a particular coin toss’s landing heads, then the law might be “All tosses of a fair coin land heads with physical probability one-half” and the additional fact the fairness of the coin in question. In what follows it is assumed that the latter facts are part of the background knowledge, and that  $t$  simply asserts some probabilistic law of nature, as suggested by the previous notation.

The most puzzling aspect of the probability coordination principle is the nature of admissibility. Lewis proposes a working definition of admissibility (he says that it is a “sufficient or almost sufficient” condition for admissibility) on which information is admissible either if it is historical—if it concerns only facts about the past up to the point where the principle is invoked—or if it is purely probabilistic, that is, if it is information about physical probabilities themselves.

The definition is problematic for two reasons. One difficulty is explicitly identified by Lewis (1980) and for many years prevented him from advancing the frequency-based theory of physical probability that he wished to give. As noted earlier, when coordinating probabilities for a given outcome, information about the future occurrence or otherwise of that outcome ought to be counted inadmissible. It turns out that frequency-based probabilities provide information of this sort. Lewis, then, has three choices. The first is to revise the working definition of admissibility so as to rule out such information, in which case information about physical probabilities will be inadmissible and the resulting probability coordination principle will be useless. The second is to stay with the working definition of admissibility, allowing the information provided by frequency-based probabilities to count as admissible by fiat. It can be shown, however, that the resulting principle—that is, Lewis’s original principle—clearly sets the wrong subjective probabilities in certain circumstances: There are certain complex facts about the future that a frequency-based probability distribution entails cannot

obtain, yet assigns a nonzero probability. If such a probability distribution is known to be the correct one, then the right subjective probability for the facts is zero, but probability coordination results in a nonzero subjective probability. The third option is to abandon probability coordination as such. Lewis takes the third way out, proposing a new kind of probability coordination principle that has the form (using the notation from earlier)  $C(e|tk) = P_i(e|t)$ . Michael Strevens (1995) points out that both Lewis's new principle and his original principle are consequences of a more general probability coordination principle according to which conditional subjective probabilities should be set equal to conditional physical probabilities. This principle yields Lewis's original principle when information about physical probability distributions is admissible and Lewis's new principle when it is not.

A different problem with Lewis's working definition of admissibility is that it makes no sense of probability coordination in deterministic systems. If one conditionalizes on the exact initial conditions of a coin toss, one ought not to set one's subjective probability for heads to the physical probability of heads, one-half, but either to zero or to one depending on whether those particular initial conditions cause the coin to land heads or tails. If a probability coordination principle is to be applied to the probability of heads, exact information about initial conditions must therefore be ruled inadmissible. Lewis's (1980) working definition of admissibility counts initial conditions, like all historical facts, as admissible.

Lewis (1980) does not regard this as a problem, since he agrees with the propensity theorists that in deterministic systems there could be only ersatz physical probabilities. Even if this is correct as a metaphysical doctrine, however, it remains a matter of fact that one coordinates one's subjective probabilities with such ersatz probabilities all the time, as when one forms expectations about the outcomes of a tossed coin. Whatever one calls it, then, there is a coordination principle for systems such as gambling devices that apparently has the same form as the genuine probability coordination principle (for a reconciliation of Lewis's account of physical probability and probability coordination in deterministic systems, see Loewer 2001).

There is clearly more work to be done elucidating the form of the probability coordination process, and in understanding admissibility in particular. A different project attempts to justify the practice of probability coordination, by giving an a priori argument that subjective probabilities should track physical probabilities, or

beliefs about such. Lewis himself says no more than that he can "see dimly" why probability coordination is rational. Howson and Urbach (1993) attempt a full-blown justification. Strevens (1999) argues that Howson and Urbach's argument appeals implicitly to a principle of indifference and goes on to make a case that there is a strong parallel between providing an a priori justification for probability coordination and providing an a priori justification for inductive inference, that is, solving the problem of induction.

A final question about the relation between epistemic and physical probability was adumbrated earlier: Why should the same formal structure be central to one's understanding of two such different things as the production of the probabilistic patterns and the nature of inductive reasoning?

**See also** Bayes, Bayes' Theorem, Bayesian Approach to Philosophy of Science; Confirmation Theory; Decision Theory; Determinism and Indeterminism; Explanation; Statistics, Foundations of.

### Bibliography

- Armstrong, David Malet. *What Is a Law of Nature?* New York: Cambridge University Press, 1983.
- Braithwaite, Richard Bevan. *Scientific Explanation: A Study of the Function of Theory, Probability, and Law in Science*. Cambridge, U.K.: Cambridge University Press, 1953.
- Carnap, Rudolf. *Logical Foundations of Probability*. Chicago: Chicago University Press, 1950.
- Daston, Lorraine. *Classical Probability in the Enlightenment*. Princeton, NJ: Princeton University Press, 1988.
- De Finetti, Bruno. "Foresight: Its Logical Laws, Its Subjective Sources." In *Studies in Subjective Probability*. 2nd ed., edited by Henry E. Kyburg Jr. and Howard E. Smokler. New York: Wiley, 1964.
- Fetzer, James H. "Dispositional Probabilities." *Boston Studies in the Philosophy of Science* 8 (1971): 473–482.
- Fine, Terrence L. *Theories of Probability: An Examination of Foundations*. New York: Academic Press, 1973.
- Giere, Ronald N. "Objective Single-Case Probabilities and the Foundation of Statistics." In *Logic, Methodology, and Philosophy of Science: Proceedings*, edited by Patrick Suppes, et al. Amsterdam, Netherlands: North Holland, 1973.
- Gillies, Donald. *Philosophical Theories of Probability*. London: Routledge, 2000.
- Hacking, Ian. *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction, and Statistical Inference*. New York: Cambridge University Press, 1975.
- Hájek, Alan. "What Conditional Probability Could Not Be." *Synthese* 137 (2003): 273–323.
- Howson, Colin, and Peter Urbach. *Scientific Reasoning: The Bayesian Approach*. 2nd ed. Chicago: Open Court, 1993.

- Jaynes, Edwin T. E. *T. Jaynes: Papers on Probability, Statistics, and Statistical Physics*, edited by R. D. Rosenkrantz. Dordrecht, Netherlands: D. Reidel, 1983.
- Jeffrey, Richard C. *The Logic of Decision*. 2nd ed. Chicago: Chicago University Press, 1983.
- Keynes, John Maynard. *A Treatise on Probability*. London: Macmillan, 1921.
- Laplace, Pierre Simon de. *A Philosophical Essay on Probabilities*. Translated by F. W. Truscott and F. L. Emory. New York: Wiley, 1902.
- Lewis, David. "Humean Supervenience Debugged." *Mind* 103 (1994): 473–490.
- Lewis, David. "A Subjectivist's Guide to Objective Chance." In *Studies in Inductive Logic and Probability*. Vol. 2, edited by Rudolf Carnap and Richard C. Jeffrey, 83–132. Berkeley: University of California Press, 1980.
- Loewer, Barry. "Determinism and Chance." *Studies in History and Philosophy of Modern Physics* 32 (2001): 609–620.
- Mellor, D. H. *The Matter of Chance*. Cambridge, U.K.: Cambridge University Press, 1971.
- Popper, Karl R. "The Propensity Interpretation of Probability." *British Journal for the Philosophy of Science* 10 (1959): 25–42.
- Ramsey, Frank Plumpton. "Truth and Probability." In *Philosophical Papers*, edited by D. H. Mellor. Cambridge, U.K.: Cambridge University Press, 1931.
- Reichenbach, Hans. *The Theory of Probability: An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability*. 2nd ed. Translated by Ernest H. Hutten and Maria Reichenbach. Berkeley: University of California Press, 1949.
- Skyrms, Brian. *Causal Necessity: A Pragmatic Investigation of the Necessity of Laws*. New Haven, CT: Yale University Press, 1980.
- Strevens, Michael. *Bigger than Chaos: Understanding Complexity through Probability*. Cambridge, MA: Harvard University Press, 2003.
- Strevens, Michael. "A Closer Look at the 'New' Principle." *British Journal for the Philosophy of Science* 46 (1995): 545–561.
- Strevens, Michael. "Inferring Probabilities from Symmetries." *Noûs* 32 (2) (1998): 231–246.
- Strevens, Michael. "Objective Probability as a Guide to the World." *Philosophical Studies* 95 (1999): 243–275.
- Suppes, Patrick, et al., eds. *Logic, Methodology, and Philosophy of Science: Proceedings*. Amsterdam, Netherlands: North Holland, 1973.
- Suppes, Patrick. "New Foundations of Objective Probability: Axioms for Propensities." In *Logic, Methodology, and Philosophy of Science: Proceedings*, edited by Patrick Suppes et al. Amsterdam, Netherlands: North Holland, 1973.
- Van Fraassen, Bas C. *Laws and Symmetry*. New York: Oxford University Press, 1989.
- Van Fraassen, Bas C. *The Scientific Image*. New York: Oxford University Press, 1980.
- Von Mises, Richard. *Probability, Statistics, and Truth*. 2nd ed. New York: Macmillan, 1957.

**Michael Strevens (2005)**

## PROCLUS

(412–485)

Proclus was born in Constantinople into a Lycean family that was still faithful to the old Hellenic religion in a society already dominated by Christianity. The talented young man forsook a career as a lawyer and decided to devote his life entirely to philosophy. After studies in Alexandria, he arrived in 431 in Athens where he joined the Platonic school of Syrianus. After the death of his venerated master, he became the leader of this school and remained in that position for almost fifty years until his death in 485. As we know from his biographer Marinus, his whole life was devoted to teaching and writing. Proclus was also a deeply devout person. In the community of the school he continued to practice with his disciples the rituals of the old Hellenic religion as well as the theurgical rituals of the Chaldeans. For Proclus, Plato was more than a philosopher intent upon the search for the truth; he was also a divinely inspired prophet showing the soul a way of salvation. Reading Plato had become more than just a scholarly exercise—it was a religious activity of paramount importance.

Proclus was convinced that the truth had been revealed by the gods in many different ways, in obscure oracles, myths and symbols. He saw himself as the interpreter whose task it was to explain the hidden significance of those religious traditions in a civilization where they were doomed to disappear. It was his ambition to prove the harmony between Plato and the other sources of divinely inspired wisdom, in particular the Chaldean Oracles and the Orphic poems. In his view, only a philosophical approach could offer the framework and rational arguments needed for this interpretation. For that reason Plato remained for him the ultimate authority in all matters, divine and human. Aristotle, on the contrary, was given only a subsidiary role, as he never developed a proper theology. His significance was restricted to matters of logic and physics.

Proclus wrote commentaries on the dialogues of Plato that were part of the curriculum of the Neoplatonic school. The course started with the reading of the *Alcibiades I* about self-knowledge, which was regarded as an introduction to philosophy, and culminated in the explanation of the two major dialogues of the Platonic corpus: the *Timaeus* about the generation of the physical world and the *Parmenides*, which was thought to offer Plato's doctrine on the first principles. Proclus also wrote a series of interpretative essays on the *Republic*. The commentaries of Proclus are masterpieces of their genre: They not only

offer a systematic interpretation of the text, but also introduce his own philosophical views and provide us with a wealth of information about the discussions within the Platonic tradition. That is particularly the case with the commentary on the *Timaeus*: Proclus defends Plato's explanation of the physical world as superior to that in Aristotle's *Physics* because only Plato discovered the ultimate (i.e., divine) causes of the physical phenomena.

Besides his work on Plato, Proclus composed a remarkable commentary on Euclid's *Elements*, the prologues of which offer a philosophical introduction to the study of the mathematical sciences. The *Hypotyposis*, or *Outline of the Astronomical Hypotheses*, is another indication of his interest in science and cosmology. Proclus also wrote short treatises on diverse subjects, such as the treatise *On the Existence of Evils*, in which he attempts to explain the existence of evil in a world proceeding from an absolutely good principle. If all agents act for some good and yet, evil occurs, it is unintended and uncaused. Evil cannot exist in its own right and no proper cause explains it. Its existence is parasitic (*para-hypostasis*), supervening upon substances and acts. This doctrine, which was adopted by a Christian author writing under the pseudonym of Pseudo-Dionysius in his celebrated treatise *On the Divine Names*, became for centuries the dominant view in philosophical and theological debates on evil.

Besides his commentaries, Proclus owes his reputation to his two great syntheses of Platonic philosophy, the *Elements of Theology* and the *Platonic Theology*. Theology is for Proclus a rational investigation into the first causes and principles of all things. The first philosophers only admitted corporeal entities, such as fire or water, as first causes. Later philosophers recognized souls as principles of life and movement and thus discovered noncorporeal being. Aristotle posited unmoved intellects above the self-moving souls and considered the first intellect to be the ultimate divine cause. Only Plato, however, recognized a cause beyond intellect, beyond being, beyond knowledge and discourse, namely the One, from which all things including matter derive their existence (*Platonic Theology* I,1). Therefore, Plato's theology is for Proclus (and for the entire Neoplatonic school since Plotinus) the accomplishment of all preceding theological speculation, since it reveals the "three principal hypostases": the One, the Intellect, the Soul.

## ELEMENTS OF THEOLOGY

In this work, the metaphysical counterpart of Euclid's *Elements*, Proclus demonstrates "in a geometrical manner" the

most fundamental theorems of the theological or metaphysical science as he understands it. The work contains 211 propositions, each of them followed by a demonstration. The first part (props. 1–112) examines the fundamental principles that govern the structure of all reality, such as the relation between the One and the many, the cause and the effect, the whole and the parts, transcendence and participation, procession and reversion, continuity and discontinuity, Being, Life and Intellect, limit and limitedness, self-movement and self-constitution, act and potency, eternity and time.

In the second part (props. 113–211) Proclus gives a survey of all degrees of reality, applying to them the general metaphysical principles he had demonstrated before. He discusses successively the gods (or "henads"), the intellects and the souls. The physical realm falls outside the scope of this theological metaphysics. The *Elements of Theology* is without doubt the most original work of Proclus, not so much because of its content (which offers the standard doctrine of the Athenian school), but because of its extraordinary attempt to develop the entire Neoplatonic metaphysics from a set of axioms. It also had a tremendous influence, in particular through the Arabic adaptation that was made in the ninth century in the circle of Al-Kindī. In the middle of the twelfth century this Arabic treatise was translated into Latin. The *Liber de Causis*, as it was named, circulated as the work of Aristotle and thus obtained a great authority in medieval Scholasticism. The systematic character of the *Elements* and its rigorous method make it for the student the best introduction to the complicated thought of Proclus.

**ONE AND MULTIPLICITY.** The *Elements* begin with the proposition that "every manifold in some way participates in unity." Without some form of unity a multitude would fall apart into an infinity of infinite things. A multitude cannot, however, be itself the unity it participates in. It is not the One, but a unified manifold, having unity as an attribute, and is therefore posterior to the One itself upon which it depends. All things, then, derive their being ultimately from the One from which they proceed.

This One must be identified with the Good, since it is the proper function of the One to hold together all things and maintain them in existence, which is also the function of the Good. For to hold a thing together and make it one is to give it its perfection and well-being, whereas dispersion is the cause of its destruction and evil. Since the Good is what unifies things and the One is what gives them perfection, the One and the Good are names

designating the absolute principle from which and toward which all things eternally proceed and return.

Having demonstrated the necessity of the One as the first principle, Proclus explains how all things in all their specificity proceed from this One. It is impossible to admit that the utmost multiplicity of the material world with its particularized bodies would proceed immediately from the first principle. Plotinus had already argued that from the One comes first the Intellect and from the Intellect proceeds the Soul, which stands itself at the beginning of time, division, and movement. For Proclus, this Plotinian understanding of the procession is unsatisfactory, in particular with regard to the second level, the Intellect, which is identical with true Being and Life. If we respect the “law of continuity,” which governs the procession of all things along the “chain of being,” we cannot admit that the Intellect (which contains already the Forms of all things) comes into existence immediately after the absolute One. There must be “mean terms” connecting the extremities. From the One comes forth Being, from Being Life, from Life the Intellect. Whereas Being is the ultimate intelligible object (*noēton*), the Intellect, which contains in its thought the paradigmatic Forms, is the properly intellectual level (*noeron*). The intermediate realm of Life is the intelligible-intellectual. In this triad of hypostases, the superior level has the most comprehensive and farthest reaching causality: for all things participate in being, but not all are living or capable of thinking. The causality of the One reaches even further than that of Being, since matter, the indeterminate substrate of the physical realm, does depend on the One, though it does not really “exist.”

**THE TRIADIC DYNAMIC STRUCTURE OF REALITY.** Many propositions in the *Elements* concern causality (hence the Latin-Arabic adaptation is appropriately called *De causis*): they are not about the physical causes, which are for Proclus only auxiliary causes, but about the “true causes,” which always transcend their effect. Whatever produces something must be superior to the effect, which owes its existence to it. If this effect has itself the power to produce, it will produce again something inferior to it, until the procession comes down to what is altogether unproductive, that is, matter. Although the effect is inferior to its cause, it is also somehow similar to that which has produced it. The effect is in a secondary manner what its cause is primarily. Insofar as the effect is similar to its cause and shares its character, it is said to “remain” in its cause without yet having its proper existence. On the superior level, it exists “causally” or “potentially” (if “potency” is understood as a productive power).

A being only acquires its proper existence (*hyparxis*) when it proceeds from its cause and becomes distinguished from it. Through the procession it becomes somehow dissimilar to the cause. Yet the procession from the cause cannot go on infinitely: the effect must also revert upon the cause from which it proceeds. Through this “return” (*epistrophē*) the effect strives to be connected again with its cause and becomes similar to it. If things have their being through procession, they attain their well-being or perfection through reversion. For the cause of their well-being can only come from the origin of their being. The final cause is thus identical with the efficient, since all things desire as ultimate end that which is the principle of their procession. As Proclus formulates it: “All that is produced by a cause both remains in it and proceeds from it” (*Elements of Theology*, § 30). “All that proceeds from something reverts upon that from which it proceeds” (§ 31). Therefore, “all that proceeds from a principle and reverts upon it has a cyclical activity” (§ 33). All beings remain in their causes, proceed from them, and return to them, in an eternal circularity, since the end is identical with the origin. Proclus finds this triadic dynamic structure on all levels of reality.

**PARTICIPATION AND NONPARTICIPATION.** When attempting to understand the relation between the Forms and the many things that are similar to them, Plato introduced the metaphor of “participation.” Participation, however, raises as many problems as it solves, as Plato shows in the aporetic discussion of the *Parmenides* (which offered ammunition for Aristotle’s subsequent criticism). The term seems to suggest that the many things sharing in the same Form take “parts” of it. How can one reconcile the transcendence of the Forms with their presence in the many things? If participation is real, the Forms must be immanent in the things sharing them and hence will be divided. But how, then, can the transcendence of the Forms be preserved? If, on the other hand, we stress the unity and the indivisibility of the Forms, we end up making participation impossible.

Proclus’s solution to this problem is the distinction between the participated and unparticipated mode of a *hypostasis*. What is participated in by the particular things cannot be the ideal Form itself, but must be a form that comes forth from it and is present in them. These immanent forms are somehow comparable to the Aristotelian Forms in matter. However, whereas Aristotle rejects the transcendent Forms as an unnecessary duplication of reality, Proclus argues that the unparticipated Forms are necessary to guarantee the universal character of the forms in matter. The participated form belongs entirely

to the particular thing sharing it. Since what inheres in one thing cannot be present in another, there is no explanation of the fact that the many things, though obtaining a proper form, have this form in common. By postulating an unparticipated Form, which exists prior to all participated forms proceeding from it, the Platonists can explain how the *eidos* is common to all that can share in it and nevertheless the same in all. As is said in proposition 23: “all that is unparticipated brings forth from itself the participated; and all the participated hypostases extend back to the unparticipated.”

The distinction between the participated and the unparticipated not only applies to the Forms, but to all levels of reality: Soul, Intellect, and even the One. Within each realm a distinction must be made between the first unparticipated term (the “monad”) and the “series” or multiplicity of beings of a similar nature coordinated with it. Thus, besides the many souls that are participated in on various levels by different bodies—the particular souls by which each human being exists as a particular animal, the souls of demons, the planetary divine souls—we must postulate the existence of the unparticipated Soul, from which a multiplicity of souls proceeds according to diverse modes of participation.

Similarly, besides the many particular intellects participated in by different divine and human souls, there must also exist an absolute unparticipated Intellect, which comprises the totality of all Forms. The many intellects proceed from this absolute Intellect and form together with it a coordinate series of a similar intellectual nature. Following the same line of reasoning, we must also posit after the One, which is absolutely transcendent and can in no way be participated by the inferior levels, a manifold series of “ones,” “units,” or “henads” consequent upon the primal One, which are participated in by the different classes of being. Those henads are not the modalities of unity acquired by beings, but self-subsisting units which remain transcendent above the beings that depend upon them. Though they are in themselves beyond being and beyond knowledge, as is the primal One, in which they remain co-united, their distinctive properties can be inferred indirectly from the different classes of beings dependent upon them. “For differences within an order of participants are determined by the distinctive properties of the principles participated in” (*Elements of Theology* § 123).

In view of the different classes of beings depending upon them, we can distinguish, for example, intelligible, intellectual, hypercosmic, or encosmic henads. Yet, insofar as they are all self-subsisting units, they remain uni-

fied in the One itself. If the One stands for the first divine cause, the different henads constitute the different classes of the gods. “For every god except the One is participable” (§ 116). With this doctrine of the henads, Proclus can defend—against Christian monotheism—both the unity and multiplicity of the divine. In his view, it is the main task of a Platonic philosopher to explain in a rational system the procession and the distinctive properties of all the classes of the gods we know through the diverse religious traditions. That is the purpose of Proclus’s last magnum opus.

## THE PLATONIC THEOLOGY

Proclus distinguishes four types of theological discourse. Divinely inspired poets use dramatic stories (talking about sexual relations, births, fights, cuttings of organs) to symbolically indicate the processions of the divine principles and their mutual relations. This mythological discourse is characteristic of the ancient Hellenic theology, as known through the Orphic poems and the works of Homer and Hesiod. In oracular discourse (in particular the *Chaldean Oracles*) prophets reveal the names and properties of the gods without resorting to the dramatic scenery of mythology. The Pythagoreans resort to mathematical analogies and similitudes (numbers, circles, spheres) to disclose the divine orders. Finally, there is scientific or dialectical theology, which investigates the divine classes and their properties using strictly rational arguments and an abstract philosophical vocabulary: one and many, being, whole and part, identity and otherness, similarity and dissimilarity.

This scientific theology has been brought to perfection by Plato in his dialogue *Parmenides*. In Proclus’s interpretation, this dialogue displays the fundamental axioms and basic concepts needed for the development of a scientific theology. In the second part of this dialogue, Parmenides examines in a dialectical exercise the hypothesis of Unity, considering the consequences following from the position of the One and from its denial, both for the One and for what is other than the One. If we start from the hypothesis of the One, only negative conclusions seem to follow: the One has no parts and is not a whole; it is not in something nor in itself; it is neither similar nor dissimilar. One cannot even say that it “is” or “is one.” In short, no names, no discourse, no knowledge of it is possible. Parmenides therefore has to restate his original hypothesis, now emphasizing the existence of the One. All attributes that were denied in the first hypothesis can be predicated of this One-that-is.

The interpretation of the different hypotheses of the *Parmenides* (of which we mention only the first two) led to a lively debate in the Neoplatonic school, as we know from Proclus's commentary. Proclus defends a theological interpretation of this dialectical discussion about the One and the Many. If the "One" stands for the first principle, the successive hypotheses of the *Parmenides* must refer to the different principles of the whole of reality. The One of the first hypothesis, of which no discourse is possible, is the absolute, unparticipated One or primal god. In the second hypothesis, Parmenides deduces, through the subsequent conclusions following from the position of the One-that-is, the different modes of unity ("henads") that are participated in by the different degrees of being. Whereas the first hypothesis leads to a negative theology, the deductions from the second hypothesis give the articulations of a positive theology. "In this dialogue proceed all the divine classes in good order from the first cause and demonstrate their mutual connection" (*Platonic Theology*, I, ch. 7).

When interpreted in this way, the *Parmenides* provides a framework in which the other discourses about the gods can be integrated and decoded: the mythological stories about Zeus and Kronos from the Hellenic and Orphic traditions, the strange divine names revealed in the *Chaldean Oracles*, the mathematical *theologumena* of the Pythagoreans, the various scattered remarks about the gods in the other dialogues of Plato. In the Renaissance, Marsilio Ficino adapted the model of Proclus's theology in an original way to integrate the revealed truth of Christianity.

## CONCLUSION

It is difficult to evaluate the originality of a thinker who, in most of his works, claims to be nothing but a faithful follower of his master Syrianus. It is Proclus, however, who put his mark on the development of the later tradition of Neoplatonism in Byzantine, Arabic, and Latin medieval thought. His huge influence—much greater than that of Plotinus—is to be explained mainly by two important indirect channels: the Christian reception of his theology by Pseudo-Dionysus and the Arabic adaptation of the *Elements* in the *Liber de Causis*. And yet it is no historical accident that Proclus gained this fame. The *diadochos* (or successor) of Plato, as he was named, has been the authoritative commentator of Plato and the great systematizer of Neoplatonic metaphysics.

**See also** *Liber de Causis*; Neoplatonism; Plato; Plotinus; Pseudo-Dionysus.

## Bibliography

### THE LIFE OF PROCLUS

*Marinus, Vie de Proclus*. Edited and translated by Henri Saffrey and Alain Segonds. Paris: Les Belles Lettres, 2003.

### ANNOTATED EDITIONS AND TRANSLATIONS

*The Elements of Theology*. 2nd ed. Edited and translated by E. R. Dodds. Oxford: Clarendon Press, 1992.

*A Commentary on the First Book of Euclid's Elements*. Translated by Glenn Morrow. Princeton, NJ: Princeton University Press, 1970.

*Proclus' Commentary on Plato's Parmenides*. Translated by Glenn Morrow and John M. Dillon. Princeton, NJ: Princeton University Press, 1987.

*Commentaire sur le Timée*. 5 vols. Translated by A. J. Festugière. Paris: Vrin, 1966–1968.

*Commentaire sur la République*. 3 vols. Translated by A. J. Festugière. Paris: Vrin, 1970.

*Théologie platonicienne*. 6 vols. Edited and translated and edited by Henri Saffrey and Leendert Westerink. Paris: Les Belles Lettres, 1968–1997.

*Commentaire sur le Premier Alcibiade*. 2 vols. Edited and translated by Alain Segonds. Paris: Les Belles Lettres, 1986. English translation by W. O'Neill (2nd ed., The Hague: Nijhoff, 1971).

*Commentaire sur le Parménide*. Vol. 1. Edited and translated by Alain Segonds and Carlos Steel. Paris: Les Belles Lettres, 2005.

*On the Existence of Evils*. Translated by Jan Opsomer and Carlos Steel. London: Duckworth, 2003.

*Proclus' Hymns*. Translated and annotated by R. M. Van den Berg. Leiden, Boston, Köln: Brill, 2001.

### BIBLIOGRAPHIES

Steel, Carlos, ed. *Proclus: Fifteen Years of Research: A Critical Bibliography (1990–2004)*. *Lustrum* 44 (2005).

For a complete survey of editions and translations, see *Plato Transformed*. Leuven: De Wulf-Mansion Centre, Katholieke Universiteit Leuven. <http://www.hiw.kuleuven.ac.be/dwmc/plato/proclus/index.htm>.

### SECONDARY LITERATURE

Siorvanes, Lucas. *Proclus. Neo-Platonic Philosophy and Science*. New Haven and London: Yale University Press, 1996.

Segonds, Alain, and Carlos Steel, eds. *Proclus et la théologie platonicienne*. Leuven-Paris: Leuven University Press—Les Belles Lettres, 2000.

*Carlos Steel (2005)*

## PRODICUS OF CEOS

Prodicus of Ceos, the Greek Sophist, was probably born before 460 BCE and was still alive at the time of the death of Socrates in 399 BCE. He traveled widely as an ambassador for Ceos and also earned a great deal of money lecturing in various Greek cities, especially in Athens. His writings are known to have dealt with physical doctrines,



with religious and moral themes, and above all with distinctions between the meanings of words usually treated as synonyms. Socrates attended a lecture by him on the last of these topics and regularly claimed to be a pupil of Prodicus in the art of synonymy (*Protagoras* 341A, *Meno* 96D).

In physics he appears to have treated the four elements of Empedocles as divine, and no doubt they formed the basis of the cosmology of Prodicus, to which Aristophanes refers in the *Birds* (1.692), although the fanciful cosmology that follows is probably not based on that of Prodicus. Prodicus further held that those natural objects and powers that are useful to human life were made the objects of cult and treated as gods by men. Inevitably, he was later classed as an atheist, but it is more likely that he offered an account of the origin of the gods that was not intended to deny their existence.

In a work titled the *Horae* (Hours) he included the since famous story “Heracles Where the Road Divides,” of which we have a fairly full summary in Book II of Xenophon’s *Memorabilia*. Vice and Virtue appear to Heracles personified as women and invite him to choose between them. Each describes what she has to offer, and Heracles chooses the arduous tasks of Virtue rather than the pleasures of Vice.

Of greater philosophic interest is the ethical relativism attributed to Prodicus in the pseudo-Platonic dialogue the *Eryxias*. There he is apparently quoted as arguing that what is good for one man is not good for another man, so that we cannot speak of anything as good *simpliciter*. On the other hand, the goodness of a thing does not depend on the goodness of the user (although some scholars have interpreted him this way). Rather, the value of a thing inheres in the thing itself in such a way that it will be good in relation to one person and not good in relation to another, according to the person and the way in which it is used.

The discussion of synonyms and the right use of words clearly involved fine distinctions of meaning between words. Many examples quoted are ethical, and a term of narrower application is commonly distinguished from one of wider application that includes in its range of meaning the meaning of the first term. The value of such distinctions is clear in rhetorical argument. But Prodicus was also eager to reject the kind of view found in Democritus, according to which there can be different names for the same thing since names are attached to things by convention only. Prodicus maintained, it would seem, that no two words have the same meaning, and in this he at least prepared the way for the search for precisely stated meanings that later fascinated Socrates and Plato.

*See also* Ethical Relativism; Sophists.

## Bibliography

### TEXTS AND TRANSLATIONS

- Diels, H., and W. Kranz, eds. *Fragmente der Vorsokratiker*. 10th ed. Vol. II. Berline, 1961.
- Sprague, R. K., ed. *The Older Sophists: A Complete Translation by Several Hands of the Fragments in die Fragmente der Vorsokratiker*, ed. by Diels-Kranz. Columbia, SC: University of South Carolina Press, 1972.

### STUDIES

- Guthrie, W. K. C. *The Sophists*. Cambridge, U.K.: Cambridge University Press, 1971.
- Kerferd, G. B. *The Sophistic Movement*. Cambridge, U.K.: Cambridge University Press, 1981.
- Kerferd, G. B. “The ‘Relativism’ of Prodicus.” *Bulletin of the John Rylands Library* 37 (1954): 249–256.

G. B. Kerferd (1967)

*Bibliography updated by Paul Woodruff (2005)*

## PROGRESS, THE IDEA OF

In broad terms a popular belief in “progress” means the rejection of an attitude that has characterized most human communities throughout history. Normally, people have believed that the future would repeat the past. When they have expected that human life was going to change, they have usually supposed that this change was going to take place suddenly and radically, by supernatural intervention. And if they have permitted themselves to hope for the improvement of the human condition, the hope has commonly been directed toward salvation from the world rather than reform of the world. By and large, historical change, when people have been aware of it at all, has been viewed as a sign of mortality and the proof of a lapse from ideal standards. Indeed, in many societies there has been a popular conviction that humankind’s condition has changed in the course of history but for the worse. Characteristically, when people have believed in a golden age, they have put that age in the past rather than the future.

In contrast, in modern Western societies change and innovation have a different place in the popular imagination. Not everyone assumes that all change is necessarily for the better, but it is widely assumed, even by conservatives, that only a society that has a general capacity to change is capable of surviving. And despite wars and depressions a large proportion of the members of Western societies have tended to expect that, short of a cataclysm, their children would live happier and better lives

than they. They have supposed that this improvement would be cumulative and continuing and that although temporary setbacks, accidents, and disasters might take place, human knowledge, power, and happiness would increase over the long run.

The emergence of this idea is the product of a variety of circumstances, such as the accumulation of an economic surplus, the increase of social mobility, and the occurrence of major inventions that have dramatically increased human power over nature. Over and above these, however, the idea of progress is peculiarly a response to the emergence of the unique social institution of organized scientific inquiry.

### HISTORY OF THE IDEA

Seeds of the faith in progress can be found in the works of the two great spokesmen for the new science, Francis Bacon and René Descartes. The fundamental elements of the idea itself were developed in the course of the so-called quarrel of the ancients and the moderns, which occupied writers and critics in the last part of the seventeenth century. At the heart of this controversy was a dispute over the authority that should be attributed to the opinions and examples left by the ancient writers. Was it the task of scholars to stand as sentinels at the gate, guarding against innovation and protecting established styles and beliefs? The controversy implicitly raised not only literary questions but the larger question of what attitude toward the past should govern the intellectual life.

In developing their position, the moderns argued that the partisans of the ancients were misled by a false analogy. They looked upon the ancients as their forefathers and therefore thought of the ancients as older and, in consequence, wiser than themselves. But just as the individual grows older and presumably wiser as time goes by, so does humanity. The so-called ancients were really the young men of humanity, and those alive today were the true ancients. They stood on the shoulders of their predecessors and could see farther; their wisdom and authority was greater than the wisdom and authority of their predecessors. This argument was developed with particular force by Bernard de Fontenelle in his *Digression sur les anciens et les modernes* (published in 1683).

The analogy between the history of humankind and the life of an individual had already been developed, however, by a number of writers. Blaise Pascal, for example, used it in drawing a belief in intellectual progress from an examination of the nature of scientific inquiry. In 1647, Pascal had published a study, *Nouvelles expériences*

*touchant le vide*, which encountered immediate objections from many scientists and philosophers, including Descartes, on the ground that it denied the time-honored truth that nature abhorred a vacuum. Pascal replied to one of his critics, Father Noel, that an appeal to inherited authority had no force where the study of physics was concerned. And in a longer essay, *Fragment d'un traité du vide*, he went on to give general reasons for moderating the respect for received authority. "The experiments which give us an understanding of nature multiply continually," he pointed out, "from whence it follows ... that not only each man advances in the sciences day by day, but that all men together make continual progress in them as the universe grows older." Pascal believed, however, that such progress took place only where the experimental methods of the sciences were relevant. In theology received authority set the final limits to inquiry, for there the object was not to add to the knowledge provided by ancient authority but only to understand as fully as possible what that authority revealed.

During the eighteenth century, however, and particularly in France an increasing number of intellectuals came to believe that the methods and spirit of science should be applied to all fields. In consequence, the idea of progress came to include a concept of social and moral progress. The cumulative improvement in human knowledge and power that had been brought about in the physical sciences could also be brought about in the organization of human society and the character of human conduct, it was asserted, if only the barriers that existed against the employment of rational methods in morals, religion, and politics could be removed. The Encyclopedists, chief among whom were Denis Diderot and Jean d'Alembert, led in the dissemination of this point of view. The most complete and moving expression of this faith in progress was the Marquis de Condorcet's *Esquisse d'un tableau historique des progrès de l'esprit humain*, written in 1793.

In the nineteenth century a new kind of historicist philosophy emerged that rejected the eighteenth-century conception of reason and the sharp dichotomy between the present and the past that had been made by believers in progress. This philosophy, best represented by G. W. F. Hegel, substituted the view that history followed its own inherent course of development and that this course of development embodied rational principles higher than those of merely human reason. Since this form of historicist philosophy identifies all conceivable changes as elements in an unfolding rational purpose, it deprives the idea of progress of definite meaning.

The more definite and combative eighteenth-century conception of progress, however, also continued to be a central theme in the thought of the nineteenth century. In one form or another, major figures of the century, such as Karl Marx, Auguste Comte, and John Stuart Mill, all propounded the doctrine. Although Marx, Comte, and Mill were influenced, each in his own way, by historicist ideas, each retained the characteristic eighteenth-century emphasis on the struggle between reason and superstition, on the movement of humankind away from theological and metaphysical modes of thought to positive or empirical habits of mind, and on the importance of extending the standards and methods of the sciences to all domains.

In the twentieth century the idea of progress continued to have adherents, particularly among American pragmatists, Marxists, and logical empiricists. For obvious historical reasons, however, advocates of the belief in progress have become steadily more modest in their claims since World War I, and since the turn of the twentieth century the idea of progress has been seized on by an increasing number of philosophers, theologians, and social critics as the prime fallacy of the tradition of liberalism and rationalism.

#### ANALYSIS OF THE IDEA

In tracing the history of the idea of progress, it is useful to distinguish between two motifs. Generally speaking, the belief in progress has been supported by an appeal to the progress of the sciences. In many cases, however, this appeal has consisted in showing that the sciences—usually some particular science—had uncovered fundamental truths that had been previously unknown and that progress would now take place if only these truths were accepted as guides to practice. Thus, progress has been said to be guaranteed if people lived by the fundamental principles disclosed by the science of economics, if they accepted the laws of historical development revealed by a scientific approach to history, or if they extended to the government of human society the Darwinian doctrine of evolution by natural selection. Progress has also been thought to be guaranteed if people could only come to recognize certain rational moral principles, such as universal natural rights. Such universal principles, though antecedent to any particular science, were nevertheless closely identified with science, for it was assumed that their validity would be apparent to anyone who could disencumber himself from the superstitions and prejudices of the past and that this process of disengagement was immensely accelerated by the advent of science. This con-

ception of the nature and conditions of progress lends itself to Utopian and Messianic interpretations of progress when understood as an ideal but to the reduction of the idea, in G. M. Young's phrase, "from an aspiration to a schedule" when associated with rigid, a priori approaches to the problem of improving the human condition.

A second motif in the theory of progress, however, has associated progress not with any particular discoveries of science or reason but with the unique, self-corrective methods of science. From this point of view the essential conditions for progress are the rejection of absolutes and fidelity to the principles of free, fallibilistic, experimental inquiry in all domains of thought and action. Even if we assume that it is valid to assert that the methods of science are universally applicable, this approach obviously imposes practical conditions for progress that are immensely difficult and perhaps impossible to realize. Accordingly, those who adopt this approach to the idea of progress can be taken to be saying only that there is a possibility of progress or, at best, a slow and uneven historical tendency that is characteristic only of societies possessing an appropriate ethic and social order and whose continuation is by no means ensured. In the past many proponents of the idea of progress undoubtedly underestimated the difficulties of domesticating within society at large the attitudes and habits of mind exemplified in scientific investigation. Nevertheless, insofar as their concept of progress depended simply on an appeal to the character of scientific procedure, they cannot be said merely to have offered a secularized version of older religious beliefs in a heavenly city, and criticisms of them for having done so, which are standard in much of the literature related to the history of the idea of progress, are a source of considerable confusion.

To be sure, the theories of progress that were developed in the eighteenth and nineteenth centuries are often based on a combination of these two motifs. In Condorcet's thought, for example, there can be found Utopian as well as realistic formulations of the idea of progress. Nevertheless, it is a mistake, on the whole, to associate the idea—particularly as it arose in eighteenth-century France—with the naive hope that human beings and human society could be made perfect. If we study the specific predictions that Condorcet made with regard to the future of humanity, for example, we find that he pointed ahead, with extraordinary prescience, to what are now such commonplace facts as the lengthening of life expectancy, social insurance, and the guarantee of equal

legal rights to all citizens. Although none of these has brought the happiness and general reasonableness that Condorcet assumed they would, it was historical realism on his part, not juvenile innocence, to make such predictions. An inability to imagine the wretchedness of the past, not a cold, unillusioned understanding of the present, lies behind the failure to appreciate why reasonable men in the eighteenth and nineteenth centuries should have been rhapsodic about the possibility of changes in the human condition that, in the light of contemporary heightened expectations, may tend to appear fairly modest.

**SCIENTIFIC PROGRESS.** What can be said with regard to the validity of the idea of progress? We must first ask what meaning can be assigned to the notion of scientific progress.

One frequent argument against the validity of the belief in scientific progress is that it contains a self-contradiction. The belief that there is scientific progress is usually attached to the argument that science is continually self-corrective. But if science never does anything but correct itself, is there any sense in speaking of scientific progress? Does not the concept of progress presuppose a fixed end or standard, and does not science, at any rate as interpreted by those who emphasize its fallibilism, deny that there can be fixed ends or standards? *Progress*, in short, appears to be a term without meaning, according to this view, unless it can be attached to metaphysical standards, such as absolute truth, whose status is antecedent to science.

This view fails, however, once it is recognized that progress can also refer to the solution of particular problems, not only to the movement toward a general and abstract goal. For example, meaning can obviously be assigned to the statement that science has made progress in determining the causes of malaria or in describing the characteristics of the other side of the moon. Such statements mean that there are now answers to questions to which there were no answers before and that these answers are in accord with the procedures of inquiry in force among competent scientific investigators. Once scientific progress is defined in terms of the solutions to particular problems, sense can also be given to the notion of cumulative scientific progress, for the general scientific capacity to solve problems has also tended to grow.

Some doubt has been thrown on these conclusions, however, by recent philosophers of science. Karl Popper, for example, argued that scientific theories and hypotheses are never genuinely confirmed but at best succeed

only in resisting successive efforts to falsify them. Since the capacity of a scientific conclusion to survive a series of such efforts does not prove that it will always be able to do so, it would seem to make no sense to speak of successful or true solutions of scientific problems. Popper's view, however, seems to involve an unnecessarily paradoxical way of stating the truism that all scientific conclusions are subject to correction in the future. The survival of a scientific conclusion despite successive efforts to overthrow it adds to the degree of reliability that may reasonably be ascribed to it. It is just as possible to describe the critical position of scientists toward accepted conclusions as efforts to extend the range and reliability of these conclusions as it is to describe it as the expression of a compulsion to destroy what has been inherited. The accumulation of increasingly well-tested and continuously powerful ideas by the sciences is an obvious fact of their history, but as seen by Popper, it seems almost an accidental by-product.

Doubt has also been thrown on the belief in scientific progress by the view that the history of science is the record of revolutions in scientific theory so radical in character that it is impossible to establish the continuity between the ideas of one generation and the ideas of a later one. If this were true, it would be impossible a fortiori to establish a concept of progress, since such a concept presupposes a measure of continuity in the sequence of events under examination. Underlying this view is the thesis that the confirmation by experiment of particular hypotheses always entails the use of a specific theoretical framework. When this theoretical framework changes, observations are simply run through a different set of conceptual categories. Accordingly, it makes little sense, it is argued, to say that the sciences have improved or extended their knowledge, for all that has happened is that one body of beliefs has been substituted for another. This point of view raises epistemological and methodological questions of great complexity, and there is no room to discuss them sufficiently here. It appears to leave out of account, however, the consideration that, for example, fundamental principles of Newtonian physics can, with appropriate modifications, be absorbed into modern physical theories. It also appears to underestimate the implications of the fact that these principles, without substantial modification, continue to provide reliable instruments for the explanation and prediction of events in large sectors of macrophysics.

**SOCIAL AND MORAL PROGRESS.** Assuming that both meaning and truth can be assigned to the idea of progress in science, what is the status of the belief in social and

moral progress? Obviously, the answer to this question depends in part on the standards employed as the touchstones of progress. However, some of the difficulties involved in stating and defending such standards can be circumvented if in this sphere we also define progress in terms of the successful solution of specific problems. Thus, there has been striking progress in the control of disease, in methods of farming, in material productivity, in the reduction of backbreaking labor, in the techniques of rapid mass communication, in the spread of literacy, and probably in the reduction of the amount of violence in everyday life.

Of course, it is theoretically possible to hold a moral code from whose standpoint one or more of these historical trends would be regarded as retrogressive rather than progressive. In fact, however, even though members of different contemporary cultures (and members of the same culture) hold widely disparate moral outlooks, there are few informed and disinterested observers, whatever their moral outlooks, who regard any of these trends, considered in themselves, as movements in the wrong direction. And most would also look upon many other historical trends that have characterized the modern world—for example, the development of more humane attitudes in penology, the abolition of slavery and serfdom, the spread of the doctrine of basic human rights—in a similarly favorable light. To this extent it is possible to speak with a measure of precision and truth of social and moral progress.

But this answer, of course, goes only part of the way. On at least two scores it is incomplete. First, it is reasonable to ask whether the gains that have been mentioned have not been bought at a cost that more than cancels them out; second, it is possible to ask how we are to vindicate the moral principles in terms of which we assess these gains as gains.

*The cost of progress.* It is not possible, of course, to give a wholly unequivocal answer to the question of the cost of progress. The notion that large-scale historical trends can be neatly categorized as good or bad belongs to eschatology, not to mature historical analysis. If the reduction of civil violence, considered in itself, is a progressive trend, contemporary mass warfare and genocide must be considered retrogressive; if rapid mass communication is a benefit to humankind, the use of the facilities of communication for totalitarian thought control is a calamity. Moreover, the successful solution of many problems often creates new and more difficult ones. The control of disease, for example, has created a serious threat of overpopulation. And by what calculus can one

measure the gains brought about, for example, by industrial innovations against the losses brought about by mass warfare or cyclical unemployment? A moral accounting system for judging even much simpler matters than these does not exist.

Nevertheless, if the span of time we measure is sufficiently long, it remains true that on the whole the physical lot of most ordinary people has considerably improved in modern societies and that this has largely been due to the application of rational techniques to the economy. The cost has been grievous, and many of the sacrifices this progress has entailed could probably have been avoided if people had employed reasonable forethought and had shown reasonable respect for the equities. Admittedly, too, it is difficult to say whether this physical progress has made individuals “happier”; indeed, it is doubly difficult to say this, for “happiness” is in part a function of what people expect, and physical progress has meant an enormous expansion of their expectations. Nevertheless, it is doubtful that most of those who put forward the view that the costs of material progress outweigh the benefits would willingly exchange places with any but the most privileged members of past societies if they actually had the chance.

Nor must we confine ourselves to a belief in purely physical or material progress. The role of fantasy, ignorance, superstition, and fanaticism in determining the world’s affairs continues to be enormous. It is doubtful, however, whether so many members of human societies, from housewives to statesmen, have ever before thought it reasonable to make decisions on the basis of carefully acquired and sifted information, and never before have societies possessed as much knowledge about themselves and their workings as they do now, shaky and scattered though that knowledge is. Only if one thinks it morally dangerous to seek reliable information before making decisions or thinks it mistaken to try to employ rational methods in the study of human affairs can he declare such long-range social trends to be anything but progressive. Indeed, the very reason that the members of an educated modern society bear a particularly heavy burden of responsibility for the emergence of doctrines such as Nazism is that they have opportunities to be informed and judicious which members of other societies did not have. In sum, although it is not possible to say in wholesale terms that there has been moral progress, it is possible to assert that the context of human behavior has changed and that the collective capacity to achieve human purposes, whether good or ill, has enormously increased. The expectations that it is reasonable to

impose on modern social arrangements are therefore justifiably higher than those that may have been reasonable in the past. In this modified but important sense it is fair to speak of moral progress.

**Justification of moral standards.** All the preceding reflections, however, obviously presuppose the validity of a secular, liberal, and rationalistic moral code. In the end, as must be obvious, objections to the idea of progress usually turn on fundamental differences in values. Whether the validity of one fundamental moral outlook as against another can be demonstratively proved is an issue that falls beyond the scope of the present article. If we assume, however, that we cannot resolve these differences in a way that will satisfy traditional standards of demonstrative certainty, there is no so-called ultimate answer to the question of whether modern society has been the scene of genuine progress.

It is possible, however, to show that a relativistic moral philosophy is perfectly compatible with a belief in progress, for it is not true that a relativistic philosophy cannot make any meaningful statements about progress because it has to grant that there are different moral standards and that all are equally valid. First, even if there is no way of proving the absolute validity of a moral outlook, there is still a way of intelligently and objectively assessing its credentials. The moral ideals that underlie the indictment of modern civilization for its excessive individualism and egalitarianism made by T. S. Eliot, for example, would require, if they were to be seriously employed as positive programs for action, the dismantling of large segments of industrial society. Since we may assume that those who put forward such criticisms would wish medical science to continue its work, for example, and would accept a world population at something like its present size, we must conclude that their announced preferences are both unrealistic and incoherent because they are incompatible with other values that they also hold. An examination of available resources, of the costs of maintaining or instituting alternative systems of values, and of the utility of these systems as guides to the resolution of definite historical problems provides a way of choosing among competing moral outlooks and makes the choice something more than a matter of personal whim or social convention.

Second, although the philosophical relativist may believe that apart from the specification of definite problems in determinate historical contexts, there is no way of showing that a moral code is valid, this does not mean that he does not himself hold any moral standards or that he is any less attached to them than an absolutist would

be. A twenty-first-century American looking at slavery in ancient Rome, for example, will regard it as a change for the better that slavery is now illegal in Western society, and he will do so whether or not he is a relativist. And to say that he might feel different if he were a Roman is irrelevant, for he is a twenty-first-century American, not a Roman, and it would be a different person with a different identity, not he, who felt different in the hypothetical circumstances. Similarly, if the standards of people in the future change, they may well disagree with us in regard to what has been progressive in history. But if these future judgments reverse present judgments, that does not bind a relativist living here and now to accept them. Nothing in his position requires him to say that progress is any historical trend that comes to be thought desirable.

**Progress as a moral standard.** As a final consideration, it is important to recognize that the idea of progress in its most important aspect is itself a regulative moral ideal, not simply a belief about history. It represents a directing principle of intellectual and social action, instructing human beings to regard all social arrangements with a critical eye and to reject any claim that any human problem has been finally solved or must be left finally unsolved. To the extent that this idea of progress is embodied in moral codes and social systems, these codes and systems will contain deliberate provision for self-reform. The idea of progress thus represents the social application of the principle that inquiry should be kept open and that no bounds can legitimately be set to the authority of such free inquiry. As such, it would appear to be an indispensable belief for a fully liberal civilization.

**See also** Alembert, Jean Le Rond d'; Bacon, Francis; Comte, Auguste; Condorcet, Marquis de; Conservatism; Descartes, René; Diderot, Denis; Eliot, Thomas Stearns; Encyclopédie; Fontenelle, Bernard Le Bovier de; Hegel, Georg Wilhelm Friedrich; Liberalism; Logical Positivism; Marx, Karl; Marxist Philosophy; Mill, John Stuart; Pascal, Blaise; Popper, Karl Raimund; Pragmatism; Rationalism; Scientific Revolutions; Utopias and Utopianism.

### **Bibliography**

Alembert, Jean d'. *Discours préliminaire de l'Encyclopédie*.

Edited by F. Picavet. Paris, 1894.

Becker, Carl L. *The Heavenly City of the Eighteenth Century Philosophers*. New Haven, CT: Yale University Press, 1932.

Bishop, Michael, and Stich, Stephen. "The Flight to Reference, or How 'Not' to Make Progress in the Philosophy of Science." *Philosophy of Science* 65 (1998): 33–49.

Bury, J. B. *The Idea of Progress*. London: Macmillan, 1920.

- Butterfield, Herbert. *The Whig Interpretation of History*. New York: Scribners, 1951.
- Campbell, James. "Arthur Lovejoy and the Progress of Philosophy." *Transactions of the Charles S. Peirce Society* 39 (2003): 617–643.
- Carrier, Martin. "How to Pile up Fundamental Truths Incessantly: On the Prospect of Reconciling Scientific Realism with Unending Progress." In *Science at Century's End: Philosophical Questions on the Progress and Limits of Science*, edited by Martin Carrier, Gerald Massey, and Laura Ruetsche. Pittsburgh, PA: University of Pittsburgh Press, 2000.
- Comte, Auguste. *Cours de philosophie positive*. Paris, 1835–1852.
- Condorcet, Marie Jean de. *Esquisse d'un tableau historique des progrès de l'esprit humain*, edited by O. H. Prior. Paris, 1933.
- Fontenelle, Bernard de. *Digression sur les anciens et les modernes*. In *Oeuvres complètes*. 5 vols. Paris, 1825. Vol. IV.
- Frankel, Charles. *The Case for Modern Man*. New York: Harper, 1956.
- Inge, W. R. *The Idea of Progress*. Oxford: Clarendon Press, 1920.
- Kitcher, Philip. "Mathematical Progress." *Revue Internationale de Philosophie* 42 (1988): 518–540.
- Maritain, Jacques. *True Humanism*. Translated by M. R. Adamson. New York, 1954.
- Mill, John Stuart. *A System of Logic*. London, 1961. Book VI, Chs. 10–11.
- Moulines, Ulises. "Is There Genuinely Scientific Progress?" In *On Comparing and Evaluating Scientific Theories*, edited by Adam Joniksz. Atlanta, GA: Rodopi, 2000.
- Niebuhr, Reinhold. *The Nature and Destiny of Man*. New York, 1951.
- Nisbet, Robert. *History of the Idea of Progress*. Revised edition. New Brunswick: Transaction Books, 1994.
- Pascal, Blaise. *Fragment d'un traité du vide*. In *Oeuvres*, edited by Léon Brunschvicg and Émile Boutroux. Paris, 1904–1914. Vol. II.
- Rousseau, Jean-Jacques. "Discourse on the Arts and Sciences." In *The Social Contract and Discourses*. Translated by G. D. H. Cole. London, 1913.
- Spencer, Herbert. *Illustrations of Universal Progress*. New York, 1881.
- Teggart, F. J., ed. *The Idea of Progress*. Rev. ed. Berkeley: University of California Press, 1949. Introduction by George H. Hildebrand.
- Todd, A. J. *Theories of Social Progress*. New York: Macmillan, 1918.
- Turgot, A. R. J. *On the Progress of the Human Mind*. Translated by McQ. de Grange. Hanover, NH, 1929.

**Charles Frankel (1967)**

*Bibliography updated by Benjamin Fiedor (2005)*

## PROJECTIVISM

"Projectivism" has its roots in David Hume's remark in the *Treatise* about the mind's "propensity to spread itself

over external objects." We sometimes speak of properties of objects where in fact the features we notice are "projections" of our internal sentiments (or other qualities of our experience). The family of metaethical views claiming that value is a projection of our conative and affective physiological states is called projectivism by Simon Blackburn (1984), and the name has stuck. Blackburn proposes that "we say that [we] *project* an attitude or habit or other commitment which is not descriptive onto the world, when we speak and think as though there were a property of things which our sayings describe, which we can reason about, know about, be wrong about, and so on" (1984, pp. 170–171). In ethics projectivism is popular because it provides an explanation of how it is that moral judgment can have the logical role that it seems to have in deciding what to do. Believing that something has some property typically provides me with a reason to act only in conjunction with the desire to promote (or oppose) the realization of that property. But believing that something is good is (or has been taken historically to be) sufficient by itself to provide a person with a reason to act. Nor is this a coincidence; it is not that we humans happen to like good things, as we happen to like to eat sugary things. Rather, it is part of the logic of judgments of goodness that they provide reasons. How can this be? Projectivists explain: the judgment that something is good is the projection of our affinity toward it, our "appetite," as Thomas Hobbes puts it.

There are three varieties of projectivism to distinguish. The most straightforward is the error theory, advanced by J. L. Mackie (1977, see also Robinson 1948), according to which our projection of value into the world is an illusion. Ordinary moral judgments presuppose an objectivity or independence of moral properties that is simply not to be had, and so they are in error. Mackie sees moral thought and language much as an atheist sees religious talk and language. The believers are not conceptually confused, but they are ontologically mistaken. The second sort of projectivism regards moral properties as Lockean "secondary qualities," not illusions, but real properties that consist in dispositions to affect human perceivers in certain ways. According to John MacDowell (1987), a leading exponent, just as we do not understand what the blueness of an object is except as the disposition to look blue to us, so we do not understand what goodness is except as the disposition to seem good to us. The projection involved in attribution of secondary qualities, including values, involves no error at all.

A third sort of projectivism is noncognitivism, or as it is more commonly called in discussions of projec-

tivism, expressivism. The expressivist holds that moral judgments do not state propositions at all but rather serve to express some noncognitive mental state of the judge. Like secondary-quality theorists, expressivists deny that there is any mistake involved in moral judgment; true, there are no moral properties, and we speak as though there are, but this “speaking as though” is just a misleading feature of the surface grammar. In fact, according to expressivists, moral judgments do not serve the same semantic function as most declarative sentences, even though they look the same.

Blackburn’s projectivist position (the most influential one of the 1980s) develops an expressivist analysis of moral language with enough logical richness and complexity to model real moral deliberation and argument. His idea is easier to make out against the background of common criticisms of expressivism. Richard Brandt (1959), among others, noted that people’s ordinary thinking about moral judgments runs contrary to expressivism. We have generally believed that normative judgments are used to state facts, that they are true or false, and when we change our moral views we come to regard our earlier views as mistaken, not merely as different. (By contrast, when one’s taste in dessert changes, one generally regards the old preference as merely different or, at worst, childish.) Brandt complained that expressivists had given no explanation of why we are so confused. Blackburn’s theory is designed to meet such objections. While maintaining an underlying expressivist semantics, he tries to show why we speak and think as though moral judgments state facts, can be true or false, and so on.

Imagine that people initially spoke about ethics in a language like English but having a quite explicitly expressivist structure. Rather than saying, “Voting for this health-care bill is morally wrong,” they said, “Boo, voting for this health-care bill!” Now imagine that these speakers valued a kind of consistency of sentiment, so that it was regarded as a confusion if someone said, “Boo, eating mammals, and hooray, eating cows!” And suppose they also believed that some moral sensibilities could never survive reflection by a rational person, so that expressing one of those sensibilities would be conclusive evidence that the speaker simply had not thought carefully about the subject. The expressivist community might “invent a predicate answering to that attitude, and treat commitments as if they were judgments, and then use all the natural devices for debating truth” (Blackburn 1984, p. 195). Since Blackburn’s theory seeks to defend realist-style reasoning without realist metaphysics, he calls it “quasi-realism.”

An important objection to Blackburn’s quasi-realism is made by Crispin Wright (1988) and Bob Hale (1990). Our moral language has a realist surface structure, and quasi-realism seeks to vindicate this structure without giving in to realist metaphysics. But if quasi-realism is successful—if every realist-sounding thing we say can be endorsed in good faith by the quasi-realist—then how will a quasi-realist be distinguishable from a full-blooded realist? As Wright puts it, Blackburn’s program confronts a dilemma: Either it does not account for all the realist logical features of moral language, in which case it fails, or it succeeds in accounting for all of them, “in which case it makes good all the things which the projectivist started out wanting to deny: that the discourse in question is genuinely assertoric, aimed at truth, and so on” (1988, p. 35).

Despite these difficulties, projectivism deserves to be taken seriously, not just in the metaphysics of value, but in other metaphysical domains as well. For example, there have been projectivists about mental states (Dennett 1987—judging that someone has intentional states is taking “the intentional stance” toward the person), causes (saying that one event caused another is projecting one’s psychological propensity to associate events of the first kind with events of the second in temporal sequence), probability (Finetti 1972—judgments of probability project one’s degree of credence into the world), and logical impossibility (Blackburn 1984—projecting a certain kind of inconceivability). With the exception of the first, all of these sorts of projectivism are plausibly attributed to Hume, who should be regarded as the prototype projectivist.

**See also** Error Theory of Ethics; Hobbes, Thomas; Hume, David; Mackie, John Leslie; Metaethics; Noncognitivism; Realism.

### **Bibliography**

- Ayer, A. J. *Language, Truth, and Logic*. Harmondsworth, U.K., 1971.
- Blackburn, S. *Essays in Quasi-Realism*. New York: Oxford University Press, 1993.
- Blackburn, S. *Spreading the Word*. Oxford: Clarendon Press, 1984.
- Brandt, R. *Ethical Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1959.
- de Finetti, B. *Probability, Induction, and Statistics*. London: Wiley, 1972.
- Dennett, D. *The Intentional Stance*. Cambridge, MA: MIT Press, 1987.
- Geach, P. T. “Ascriptivism.” *Philosophical Review* 69 (1960): 221–225.



- Gibbard, A. *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press, 1990.
- Hale, B. "Can There Be a Logic of Attitudes?" In *Reality, Representation, and Projection*, edited by J. Haldane. New York: Oxford University Press, 1990.
- Harman, G. *The Nature of Morality*. New York: Oxford University Press, 1977.
- MacDowell, J. *Projection and Truth in Ethics*. Lawrence, KS, 1987.
- Mackie, J. L. *Ethics: Inventing Right and Wrong*. Harmondsworth, U.K.: Penguin, 1977.
- Robinson, R. "The Emotivist Theory of Ethics." *Proceedings of the Aristotelian Society* 22 (1948): supplement, 79–106.
- Smart, J. J. C. *Ethics, Persuasion, and Truth*. London: Routledge and Kegan Paul, 1984.
- Stevenson, C. L. *Facts and Values*. New Haven, CT: Yale University Press, 1963.
- Wright, C. "Realism, Anti-Realism, Irrealism, Quasi-Realism." In *Realism and Antirealism*, edited by P. French, T. Uehling, and H. Wettstein. *Midwest Studies in Philosophy*, Vol. 12. Minneapolis: University of Minnesota Press, 1988.

*James Dreier (1996)*

## PROMISES

Promising is a device for obligating oneself. In a culture in which promising is available, people have a normative power that they would lack in the absence of this institution. By exercising this power—standardly through the utterance of a linguistic formula—one can bring about changes in the expectations of others in ways that enhance one's ability to pursue their goals and that foster relations of familiarity and trust.

The existence of a normative power of this kind seems philosophically puzzling: How can the utterance of a linguistic formula cause a change in the normative relations that obtain in the world? In response to this question, it might help to situate promising within the general theory of speech acts, noting that it is one of a range of illocutionary acts that may be performed with words (to be set aside such acts as asserting or commanding). In addition, one might point to the conventional aspect of promising, observing that it is a contingent social practice, sustained by interlocking sets of human dispositions, expectations, and sanctions, which enables people to coordinate their behavior in ways that promote the common good. An analogy might be to the institution of contract in the law.

But there is a normative complexity to promises that these remarks fail to capture. Agents who make a promise incur a distinctively moral obligation, opening themselves to corresponding moral complaint if they should

fail to do what they have promised. Moreover this moral dimension seems crucial to the ordinary operation of promises. Promises serve to assure the promisee that something the promisee values will in fact take place—this is one way in which they differ from threats. But they achieve this effect through the promiser's implicit acknowledgement of the moral obligation that is brought into existence by the act of promising itself. Promisers give promisees to understand that they have a distinctively moral reason to do what has been promised, a reason that is strong enough to lead to performance even in the absence of independent reasons for so acting. This in turn grounds the promisee's assurance that the promised performance will take place. What accounts for the moral obligation that thus figures at the center of promissory interactions?

One answer to this question stresses the value to the agent of the normative power involved in promising. The ability to obligate oneself is a great advantage when it comes to pursuing one's projects and developing interpersonal relationships of depth and commitment. One would potentially deprive oneself of this advantageous capacity if one failed to do what one had promised, insofar as people would be less inclined to take one's promissory acts seriously. It may be doubted, however, whether this approach provides a complete account of the moral obligations brought into existence through promising. One issue is the directionality of promissory obligations. The normative powers approach focuses on the moral importance to the promiser of the ability to obligate oneself in this way. But when one fails to do what they have promised, the moral objection to their conduct turns primarily on the effects of their behavior on others.

This dimension of promissory obligations is central to a second approach, the practice view. On this view, the moral wrong involved in promise-breaking derives from the nature of promising as a valuable convention. This basic idea might be developed in a variety of ways, depending on the more general moral theory one favors. Thus, utilitarians invoke the duty to promote the impartial good, arguing that it is a violation of that duty to act in ways that undermine a highly beneficial social practice such as promising. Other theorists appeal to the idea of fairness, contending that it would be unfair to fail to do one's part to sustain a beneficial practice that one has profited from oneself. The moral duty that promising brings into existence is thus traced to fundamental social duties, in accordance with moral principles of utility or fairness.

There are two potential problems with this approach, however. First, it still does not capture the specific directionality of the moral duty involved in promissory acts. On the practice view, everyone who potentially benefits from the useful convention of promising could equally be said to be wronged when a person breaks a promise. Intuitively, however, it appears that the promisee, in particular, has a privileged ground for moral complaint. Second, it would seem possible to wrong another person in precisely the same way without exploiting a social practice such as promising. Thus, even in the absence of a promise A might deliberately lead B to believe that A will do X, where X is something A knows B wants A to do. Under these circumstances, A's failure to do X would appear to wrong B in just the same way a broken promise would have done. Yet this wrong cannot be explained by appeal to more general duties to sustain beneficial practices.

A third approach, the fidelity view, holds that promissory obligations derive from more general duties not to disappoint the expectations one has deliberately raised in others. This approach accounts well for the specific directionality of promissory obligations, and it does so in a way that explains the similarities between breaking a promise and other cases of dashed expectations. But the fidelity view encounters a different problem. It holds that the moral duty to keep one's promise is in place only when the promisee has come to expect that the promiser will perform. But as was seen above, in the promising case this kind of expectation is supposed to derive from the promiser's acknowledgement of the moral obligation to perform. There is thus a potential circularity in the interpretation of promissory interactions that is suggested by the fidelity view.

Much of the philosophical interest of promises derives from their normative complexity. An account that is adequate to this complexity might need to draw on several of the strategies sketched above, in a kind of hybrid approach. For instance, the practice view might explain how the act of promising brings into existence an initial moral obligation that is independent of the promisee's expectations. Perhaps it is the promiser's acknowledgement of this practice-based obligation that generates a corresponding expectation in the promisee. Once such an expectation is in place, the fidelity view could explain why promisers incur a further and specifically directional obligation to perform. Finally, the normative powers approach illuminates the value of the social practice of promising, highlighting the advantages gained from having the ability to obligate oneself through promissory acts.

*See also* Deontological Ethics; Moral Rules and Principles.

### Bibliography

- Anscombe, G. E. M. "Rules, Rights, and Promises." In *Ethics, Religion, Politics: Collected Philosophical Papers*. Minneapolis: University of Minnesota Press, 1981.
- Fried, Charles. *Contract as Promise*. Cambridge, MA: Harvard University Press, 1981.
- Kolodny, Niko and R. Jay Wallace. "Promises and Practices Revisited." *Philosophy and Public Affairs* 31 (2003): 119–154.
- Rawls, John. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1971.
- Rawls, John. "Two Concepts of Rules." *The Philosophical Review* 64 (1955): 3–32.
- Raz, Joseph. "Promises and Obligations." In *Law, Morality, and Society. Essays in Honour of H. L. A. Hart*, edited by P. M. S. Hacker and Joseph Raz. Oxford: Clarendon Press, 1977.
- Scanlon, T. M. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press, 1999.
- Searle, John R. *Speech Acts. An Essay in the Philosophy of Language*. Cambridge, U.K.: Cambridge University Press, 1969.
- Thomson, Judith. *The Realm of Rights*. Cambridge, MA: Harvard University Press, 1990.
- Warnock, G. M. *The Object of Morality*. London: Methuen, 1971.

R. Jay Wallace (2005)

## PROOF THEORY

The background to the development of "proof theory" since 1960 is contained in the entry "Mathematics, Foundations of." Briefly, Hilbert's program (HP), inaugurated in the 1920s, aimed to secure the foundations of mathematics by giving finitary consistency proofs of formal systems such as for number theory, analysis, and set theory, in which informal mathematics can be represented directly. These systems are based on classical logic and implicitly or explicitly depend on the assumption of "completed infinite" totalities. Consistency of a system S (containing a modicum of elementary number theory) is sufficient to ensure that any finitarily meaningful statement about the natural numbers that is provable in S is correct under the intended interpretation. Thus, in David Hilbert's view, consistency of S would serve to eliminate the "completed infinite" in favor of the "potential infinite" and thus secure the body of mathematics represented in S. Hilbert established the subject of proof theory as a technical part of mathematical logic by means of which his program was to be carried out; its methods are described below.

In 1931 Kurt Gödel's second incompleteness theorem raised a *prima facie* obstacle to HP for the system Z of elementary number theory (also called Peano arithmetic—PA) since all previously recognized forms of finitary reasoning could be formalized within it. In any case Hilbert's program could not possibly succeed for any system such as set theory in which *all* finitary notions and reasoning could unquestionably be formalized. These obstacles led workers in proof theory to modify HP in two ways. The first was to seek reductions of various formal systems S to more constructive systems S'. The second was to shift the aims from foundational ones to more mathematical ones. Examples of the first modification are the reductions of PA to intuitionistic arithmetic HA and Gentzen's consistency proof of PA by finitary reasoning coupled with quantifier-free transfinite induction up to the ordinal  $\epsilon_0$ ,  $TI(\epsilon_0)$ , both obtained in the 1930s. The second modification of proof theory was promoted especially by Georg Kreisel starting in the early 1950s; he showed how constructive mathematical information could be extracted from nonconstructive proofs in number theory. The pursuit of proof theory along the first of these lines has come to be called relativized Hilbert program or reductive proof theory, while that along the second line is sometimes called the program of unwinding proofs or, perhaps better, extractive proof theory. In recent years there have been a number of applications of the latter both in mathematics and in theoretical computer science. Keeping the philosophical relevance and limitations of space in mind, the following account is devoted entirely to developments in reductive proof theory, though the two sides of the subject often go hand in hand.

## METHODS OF FINITARY PROOF THEORY

Hilbert introduced a special formalism called the epsilon calculus to carry out his program (the nomenclature is related neither to the ordinal  $\epsilon_0$  nor to the membership symbol in set theory), and he proposed a particular substitution method for that calculus. Following Hilbert's suggestions, Wilhelm Ackermann and John von Neumann obtained the first significant results in finitary proof theory in the 1920s. Then, in 1930, another result of the same character for more usual logical formalisms was obtained by Jacques Herbrand, but there were troublesome aspects of his work. In 1934 Gerhard Gentzen introduced new systems, the so-called sequent calculi, to provide a very clear and technically manageable vehicle for proof theory, and reobtained Herbrand's fundamen-

tal theorem via his cut-elimination theorem. Roughly speaking, the latter tells us that every proof of a statement in quantificational logic can be normalized to a direct proof in which there are no detours ("cuts") at any stage via formulas of a complexity higher than what appears at later stages. Sequents have the form  $\Gamma \rightarrow \Delta$ , where  $\Gamma$  and  $\Delta$  are finite sequences of formulas (possibly empty).  $\Gamma \rightarrow \Delta$  is derivable in Gentzen's calculus LK just in case the formula  $A \supset B$  is derivable in one of the usual calculi for classical predicate logic, where A is the conjunction of formulas in  $\Gamma$  and B is the disjunction of those in  $\Delta$ .

## INTRODUCTION OF INFINITARY METHODS TO PROOF THEORY

Gentzen's theorem as it stood could not be used to establish the consistency of PA, where the scheme of induction resists a purely logical treatment, and for this reason he was forced to employ a partial cut-elimination argument whose termination was guaranteed by the principle  $TI(\epsilon_0)$ . Beginning in the 1950s, Paul Lorenzen and then, much more extensively, Kurt Schütte began to employ certain infinitary extensions of Gentzen's calculi (cf. Schütte, 1960, 1977). This was done first of all for elementary number theory by replacing the usual rule of universal generalization by the so-called  $\omega$ -rule, in the form: from  $\Gamma \rightarrow \Delta, A(n)$  for each  $n = 0, 1, 2, \dots$ , infer  $\Gamma \rightarrow \Delta, (x)A(x)$ . Now derivations are well-founded trees (whose tips are the axioms  $A \rightarrow A$ ), and each such is assigned an ordinal as length in a natural way. For this calculus  $LK_\omega$ , one has a full cut-elimination theorem, and every derivation of a statement in PA can be transformed into a cut-free derivation of the same in  $LK_\omega$  whose length is less than  $\epsilon_0$ . Though infinite, the derivation trees involved are recursive and can be described finitarily, to yield another consistency proof of PA by  $TI(\epsilon_0)$ . Schütte extended these methods to systems  $RA_\alpha$  of ramified analysis ( $\alpha$  an ordinal) in which existence of sets is posited at finite and transfinite levels up to  $\alpha$ , referring at each stage only to sets introduced at lower levels. Using a suitable extension of  $LK_\omega$  to  $RA_\omega$ , Schütte obtained cut-elimination theorems giving natural ordinal bounds for cut-free derivations in terms of the so-called Veblen hierarchy of ordinal functions. In 1963 he and Solomon Feferman independently used this to characterize (in that hierarchy) the ordinal of predicative analysis, defined as the first  $\alpha$  for which  $TI(\alpha)$  cannot be justified in a system  $RA_\beta$  for  $\beta < \alpha$ . William Tait (1968) obtained a uniform treatment of arithmetic, ramified analysis, and related unramified systems by means of the cut-elimination theorem for LK extended to a language with formulas built

by countably infinite conjunctions (with the other connectives as usual). Here the appropriate new rule of inference is: from  $\Gamma \rightarrow \Delta, A_n$ , for each  $n = 0, 1, 2, \dots$ , infer  $\Gamma \rightarrow \Delta, A$ , where  $A$  is the conjunction of all the  $A_n$ 's.

Brief mention should also be made of the extensions of the other methods of proof theory mentioned above, concentrating on elimination of quantifiers rather than cut elimination. In the 1960s Burton Dreben and his students corrected and extended the Herbrand approach (cf. Dreben and Denton, 1970). Tait (1965) made useful conceptual reformulations of Hilbert's substitution method; a number of applications of this method to subsystems of analysis have been obtained in the 1990s by Grigori Mints (1994). Another approach stems from Gödel's functional interpretation, first presented in a lecture in 1941 but not published until 1958 in the journal *Dialectica*; besides the advances with this made by Clifford Spector in 1962, more recently there have been a number of further applications both to subsystems of arithmetic and to subsystems of analysis (cf. Feferman 1993). Finally, mention should be made of the work of Dag Prawitz (1965) on systems of natural deduction, which had also been introduced by Gentzen in 1934 but not further pursued by him; for these a process of normalization takes the place of cut elimination. While each of these other methods has its distinctive merits and advantages, it is the methods of sequent calculi in various finitary and infinitary forms that have received the most widespread use.

## PROOF THEORY OF IMPREDICATIVE SYSTEMS

The proof theory of impredicative systems of analysis was initiated by Gaisi Takeuti in the 1960s. He used partial cut-elimination results and established termination by reference to certain well-founded systems of ordinal diagrams (cf. Takeuti 1987). In 1972 William Howard determined the ordinal of a system  $ID_1$  of one arithmetical inductive definition, in the so-called Bachmann hierarchy of ordinal functions; the novel aspect of this was that it makes use of a name for the first uncountable ordinal in order to produce the countable (and in fact recursive) ordinal of  $ID_1$ . In a series of contributions by Harvey Friedman, Tait, Feferman, Wolfram Pohlers, Wilfried Buchholz, and Wilfried Sieg stretching from 1967 into the 1980s, the proof theory of systems of iterated inductive definitions  $ID_\alpha$  and related impredicative subsystems of analysis was advanced substantially. The proof-theoretic ordinals of the  $ID_\alpha$  were established by Pohlers in terms of higher Bachmann ordinal function systems (cf. Buchholz et al. 1981). The methods here use cut-elimination

arguments for extensions of LK involving formulas built by countably and uncountably long conjunctions. In addition, novel "collapsing" arguments are employed to show how to collapse suitable uncountably long derivations to countable ones in order to obtain the countable (again recursive) ordinal bounds for these systems. An alternative functorial approach to the treatment of iterated inductive definitions was pioneered by Jean-Yves Girard (1985).

In 1982 Gerhard Jäger initiated the use of the so-called admissible fragments of Zermelo-Fraenkel set theory as an illuminating tool in the proof theory of predicatively reducible systems (cf. Jäger 1986). This was extended by Jäger and Pohlers (1982) to yield the proof-theoretical ordinal of a strong impredicative system of analysis; that makes *prima facie* use of the name of the first (recursively) inaccessible ordinal. Michael Rathjen (1994) has gone beyond this to measure the ordinals of much stronger systems of analysis and set theory in terms of systems of recursive ordinal notations involving the names of very large (recursively) inaccessible ordinals, analogous to the so-called large cardinals in set theory.

## SIGNIFICANCE OF THE WORK FOR HP AND REDUCTIVE PROOF THEORY

Ironically for the starting point with Hilbert's aims to eliminate the "completed infinite" from the foundations of mathematics, these developments have required the use of highly infinitary concepts and objects to explain the proof-theoretical transformations involved in an understandable way. It is true that in the end these can be explained away in terms of transfinite induction applied to suitable recursive ordinal notation systems. Even so, one finds few who believe that one's confidence in the consistency of the systems of analysis and set theory that have been dealt with so far has been increased as a result of this body of work. However, while the intrinsic significance of the determination of the proof-theoretic ordinals of such systems has not been established, that work can still serve behind the scenes as a tool in reductive proof theory. It is argued in Feferman (1988) that one has obtained thereby foundationally significant reductions, for example of various (*prima facie*) infinitary systems to finitary ones, impredicative to predicative ones, and nonconstructive to constructive ones. With a field that is still evolving at the time of writing, it is premature to try to arrive at more lasting judgments of its permanent value.

**See also** Gödel, Kurt; Gödel's Incompleteness Theorems; Hilbert, David; Logic, History of; Mathematics, Foundations of; Neumann, John von; Set Theory.

### Bibliography

- Buchholz, W., S. Feferman, W. Pohlers, and W. Sieg. *Iterated Inductive Definitions and Subsystems of Analysis: Recent Proof-Theoretical Studies*. Lecture Notes in Mathematics 897. New York: Springer-Verlag, 1981.
- Buchholz, W., and K. Schütte. *Proof Theory of Impredicative Systems*. Naples, 1988.
- Dreben, B., and J. Denton. "Herbrand-Style Consistency Proofs." In *Intuitionism and Proof Theory*, edited by J. Myhill, A. Kino, and R. E. Vesley, 419–433. Amsterdam: North-Holland, 1970.
- Feferman, S. "Gödel's *Dialectica* Interpretation and Its Two-Way Stretch." In *Computational Logic and Proof Theory*, edited by G. Gottlob et al. Lecture Notes in Computer Science 713. New York: Springer-Verlag, 1993.
- Feferman, S. "Hilbert's Program Relativized: Proof-Theoretical and Foundational Reductions." *Journal of Symbolic Logic* 53 (1988): 364–384.
- Gentzen, G. *The Collected Papers of Gerhard Gentzen*, edited by M. Szabo. Amsterdam: North-Holland, 1969.
- Girard, J.-Y. "Introduction to  $\Pi^1_2$  Logic." *Synthese* 62 (1985).
- Girard, J.-Y. *Proof Theory and Logical Complexity*. Naples: Bibliopolis, 1987.
- Hilbert, D., and P. Bernays. *Grundlagen der Mathematik*. Vols. 1–2, 2nd ed. Berlin: Springer, 1968–1970.
- Howard, W. "A System of Abstract Constructive Ordinals." *Journal of Symbolic Logic* 37 (1972): 355–374.
- Jäger, G. *Theories for Admissible Sets: A Unifying Approach to Proof Theory*. Naples: Bibliopolis, 1986.
- Jäger, G., and W. Pohlers. "Eine beweistheoretische Untersuchung von  $(\Delta^1_2\text{-CA})+(\text{BI})$  und verwandter Systeme." *Sitzungsber. Bayerische Akad. Wissenschaft Math. Nat. Klasse* (1982): 1–28.
- Kreisel, G. "A Survey of Proof Theory." *Journal of Symbolic Logic* 33 (1965): 321–388.
- Mints, G. E. "Gentzen-Type Systems and Hilbert's Epsilon Substitution Method. I." In *Logic, Methodology, and Philosophy of Science IX*, edited by D. Prawitz et al. Amsterdam and New York: Elsevier, 1994.
- Mints, G. E. *Selected Papers in Proof Theory*. Naples: Bibliopolis, 1992.
- Myhill, J., A. Kino, and R. E. Vesley, eds. *Intuitionism and Proof Theory*. Amsterdam: North-Holland, 1970.
- Pohlers, W. "Contributions of the Schütte School in Munich to Proof Theory." In *Proof Theory*, edited by G. Takeuti. Amsterdam: North-Holland, 1987.
- Pohlers, W. *Proof Theory: An Introduction*. Lecture Notes in Mathematics 1407. New York: Springer-Verlag, 1989.
- Prawitz, D. *Natural Deduction*. Stockholm: Almqvist and Wiksell, 1965.
- Rathjen, M. "Admissible Proof Theory and Beyond." In *Logic, Methodology, and Philosophy of Science IX*, edited by D. Prawitz et al. Amsterdam and New York: Elsevier, 1994.
- Schütte, K. *Beweistheorie*. Berlin: Springer, 1960.
- Schütte, K. *Proof Theory*. Berlin: Springer-Verlag, 1977.

Sieg, W. "Hilbert's Program Sixty Years Later." *Journal of Symbolic Logic* 53 (1988): 338–348.

Simpson, S. G. "Partial Realizations of Hilbert's Program." *Journal of Symbolic Logic* 53 (1988): 349–363.

Tait, W. "Normal Derivability in Classical Logic." In *The Syntax and Semantics of Infinitary Languages*, edited by J. Barwise. Lecture Notes in Mathematics 72. New York: Springer-Verlag, 1968.

Tait, W. "The Substitution Method." *Journal of Symbolic Logic* 30 (1965): 175–192.

Takeuti, G. *Proof Theory*. 2nd ed. Amsterdam: North-Holland, 1987.

Solomon Feferman (1996)

## PROPER NAMES AND DESCRIPTIONS

A *singular term* is an expression whose semantic function, when used in a particular context, is to *refer to* (denote, designate)—that is, to stand for—a single thing. A *definite description* is a singular noun phrase beginning with the definite article "the" or with a possessive noun or pronoun, as "the author of *Waverley*" and "my brilliant career." Proper names, such as "Shakespeare" and "London," are generally classified along with definite descriptions, individual variables, pronouns, and some other indexicals as singular terms. A French speaker who utters the words "*Londres est jolie*" asserts the same thing as an English speaker uttering "London is pretty." The thing asserted is a *proposition*, that London is pretty. The fundamental semantic role of a declarative sentence is to *express* (or to *contain*) a proposition (*q.v.*), which is the *semantic content* of the sentence. The proposition that Sir Walter Scott is ingenious has some component in common with the proposition that Scott is ingenuous, because both of these are directly about Scott, and some other component again in common with the proposition that Shakespeare is ingenious. These two proposition components are separately correlated with the proper name "Scott" and the predicate "is ingenious." The proposition component semantically correlated with an expression is the expression's *semantic content*. The principal philosophical controversy regarding proper names (and other singular terms) concerns the question: What are their semantic contents? The theories of John Stuart Mill (1806–1873), Bertrand Russell (1872–1970), and Gottlob Frege (1848–1925) provide rival answers.

## 1. THE NAIVE THEORY AND THE MILLIAN THEORY

One natural theory of semantic content is the naive theory, whose main theses are: (i) the semantic content of any singular term, as used in particular context, is its referent (bearer; the individual referred to); (ii) any semantically contentful expression refers to its semantic content; and (iii) the proposition semantically contained in a sentence is a complex, structured entity whose constituents are the semantic contents of expressions making up the sentence, typically the simple (noncompound) component expressions. (The theory may allow particular sorts of exceptions, as for example those generated by the use of quotation marks.) On the naive theory the proposition contained in “Shakespeare is ingenious” is a *singular proposition*—composed partly of things such as properties, relations, and concepts, and partly of the very individual(s) the proposition is about. By contrast, a (*purely*) *general proposition* is made up exclusively in a certain way of the former sorts of entities. On the naive theory, semantic content and reference collapse into one.

Definite descriptions pose a difficulty for the naive theory because they contain proper parts with semantic content. In *A System of Logic* (1893), Mill proffered a variant of the naive theory on which the proposition contained in “The author of *Waverley* is ingenious” is composed of something involving the attribute of authorship of *Waverley* in place of Scott himself. Mill distinguished between *denotation* (referent) and *connotation*. A general term (“concrete general name”) was said by Mill to “denote” the class of individuals to which the term applies. Mill used the term “connotation” for a semantic content consisting of attributes or properties. General terms were held to have both denotation and connotation. According to Mill, definite descriptions also have both connotation and (typically) denotation, whereas proper names have only denotation. Mill’s theory strongly suggests a systematic modification of the naive theory. The central theses of the Millian theory are: (i) the semantic content of any *simple* (noncompound) singular term is its referent; (ii) any expression refers to its extension; and (iii) the semantic content of a typical contentful compound expression (e.g., a definite description) is a composite entity whose constituents are the semantic contents of expressions making up the compound expression, typically the simple component expressions. (Mill’s actual theory was somewhat more complex, but also somewhat less plausible.)

## 2. THE PUZZLES

The naive and the Millian theories give rise to philosophical puzzles concerning substitution and nonreferring names. Frege’s puzzle arises from certain sentences, especially identity sentences. The sentence “Hesperus is Phosphorus” (or “The Evening Star is The Morning Star”), by contrast with “Hesperus is Hesperus,” is informative. Its semantic content apparently extends knowledge. It is also a posteriori and synthetic. Yet according to both the naive theory and the Millian theory, the semantic contents of both sentences are composed of the same components, evidently in precisely the same way. Those theories thus ascribe the same semantic content to both sentences. In his early work, *Begriffsschrift* (1972 [1879], §8), Frege proposed solving this puzzle by reading the predicate for numerical identity as covertly metalinguistic: It was held that “Hesperus is Phosphorus” contains a substantive proposition concerning the names “Hesperus” and “Phosphorus,” to the effect that they are co-referential. There are serious difficulties with this account, however, and Frege came to reject it. Most significantly, the account fails to solve the general problem of which “Hesperus is Phosphorus” is a special case. Unless the theory is part of a more sweeping proposal concerning all expressions and not just that of identity predicates, there is no explanation for the analogous difference in epistemic and semantic status between “Hesperus is a planet if Phosphorus is” (synthetic *a posteriori*) and “Hesperus is a planet if Hesperus is” (analytic *a priori*).

A second puzzle is the apparent failure of substitution in special contexts, especially those of propositional attitude. Jones may sincerely and reflectively assent to “Hesperus appears in the evening sky” and sincerely and reflectively dissent from “Phosphorus appears in the evening sky,” even while fully grasping their semantic content. This appears to violate the classical logical rule of Leibniz’s law, or the substitutivity of equality. Both the naive theory and the Millian theory treat “Jones believes that Hesperus appears in the evening” and its substitution instance “Jones believes that Phosphorus appears in the evening” as having the same content, and therefore also the same truth-value.

A further nest of problems concerns sentences involving nonreferring proper names. The sentence “Sherlock Holmes is addicted to cocaine” clearly has content. Yet on both the naive theory and the Millian theory, the semantic content of any sentence will lack a necessary component if any contained name lacks a referent. It is evident, moreover, that this sentence (taken as a statement of real fact, rather than as a statement made from

within the fiction) cannot be counted literally true. But, it seems, neither can its negation—“Sherlock Holmes is not addicted to cocaine”—be truly uttered. This seems to violate the classical law of excluded middle. These puzzles are especially pressing with regard to negative existentials, such as “Sherlock Holmes does not exist.” This sentence is true if and only if “Sherlock Holmes exists” is false, and therefore, it would seem, if and only if the referent of “Sherlock Holmes” lacks existence. Yet the negative existential itself implies that the name does not so much as have a referent. How, then, can it be true? Indeed, how can it have any content at all?

### 3. RUSSELL’S THEORY OF DESCRIPTIONS

Russell’s semantic theory (post-1904) is a supplement to the naive theory. Russell employed *propositional functions* in lieu of attributes. A propositional function assigns to any objects in its domain a singular proposition concerning those objects. Russell’s general theory of descriptions, or of what he called “denoting phrases,” consisting of a noun phrase preceded by a determiner such as “all” or “some,” assigns a content to sentences in which they figure while denying that the determiner phrases themselves are meaningful units. The theory analyzes sentences of both the Aristotelian *A* form, “ $\Pi(\text{all } S)$ ” (e.g., “All millionaires are wealthy”), and the *I* form, “ $\Pi(\text{some } S)$ ,” where  $\Pi$  is a monadic predicate. (More generally,  $\Pi$  may be the result of filling all but one of the argument positions of an  $n$ -adic predicate,  $n \geq 1$ .) The *A* form is analyzed as “For everything  $x$ , if  $x$  is a  $S$ , then  $\Pi(x)$ ”—more colloquially as, “Everything is such that: if it is a  $S$ , then  $\Pi(\text{it})$ ” (“Everything is, if a millionaire, then wealthy”). The complex predicate “is such that: if it is a  $S$ , then  $\Pi(\text{it})$ ” stands for a certain propositional function, whereas the quantifier “everything” stands for a higher-level propositional function, which assigns to any first-level propositional function,  $F$ , the proposition that  $F$  is “always true”—that is, the proposition that  $F$  yields a true proposition for each and every argument.

Russell analyzed “ $\Pi(\text{some } S)$ ” as “Something is such that: it is a  $S$  and  $\Pi(\text{it})$ ”—wherein the complex predicate “is such that: it is a  $S$  and  $\Pi(\text{it})$ ” stands for a certain propositional function said to be “sometimes true”—that is, to yield a true proposition for at least one argument. An English phrase of the form “all  $S$ ” thus corresponds to the incomplete string, “everything is such that: if it is a  $S$ , then it ...,” and a phrase of the form “some  $S$ ” corresponds to the incomplete string, “something is such that: it is a  $S$  and it....” Russell called phrases of either form

*incomplete symbols*. The sentences in which such phrases figure have content, though the phrase, in and of itself, does not contribute a proposition-component to the proposition expressed. As Russell put it in “On Denoting,” “denoting phrases have no meaning in isolation.”

The introduction of a quantifier (“everything,” “something”) into the analysis gives rise to ambiguities analogous to that of “every boy kissed a girl” when the simple Aristotelian sentential form occurs within the scope of a governing operator, such as “not,” “necessarily,” or “Jones believes.” Thus, on Russell’s general theory of descriptions, a sentence of the form “not  $\Pi(\text{all } S)$ ” (e.g., “All millionaires are not wealthy”) may be analyzed by giving the indefinite description “all  $S$ ” *primary occurrence* (over “not”), yielding: “Everything is such that: if it is a  $S$ , then not- $\Pi(\text{it})$ .” This reading is equivalent to the Aristotelian *E* form, “ $\Pi(\text{no } S)$ .” Alternatively, and nonequivalently, “not  $\Pi(\text{all } S)$ ” may be analyzed by giving the phrase “all  $S$ ” *secondary occurrence*, yielding the reading, “Not everything is such that: if it is a  $S$ , then  $\Pi(\text{it})$ .” (The latter analysis—equivalent to the Aristotelian *E* form—is obtained by letting the negation in “not  $\Pi(\text{all } S)$ ” govern the entire *A* form, not just its predicate  $\Pi$ .) Similarly, “Jones believes  $\Pi(\text{some } S)$ ” may be analyzed as “Something is such that: it is  $S$  and Jones believes that  $\Pi(\text{it})$ ” (primary occurrence), or alternatively, and nonequivalently, as “Jones believes: that  $\Pi(\text{some } S)$ ” (secondary).

In most cases, only one of the two readings is plausibly intended (as with “Jones believes some husbands are bachelors”). If the simple Aristotelian *A* or *I* form occurs with two or more governing operators, the number of readings is compounded. For example, “Jones believes some millionaires are not wealthy” may be analyzed alternatively, and nonequivalently, as: (i) “Someone is a millionaire and Jones believes he/she is not wealthy” (*wide scope*); (ii) “Jones believes: that someone is both a millionaire and not wealthy” (*intermediate scope*); or (iii) “Jones believes: that no one is both a millionaire and wealthy” (*narrow scope*).

The central tenet of Russell’s theory of definite descriptions is that a description such as “the author of *Waverley*” (used in the sense of “the sole author of *Waverley*”) is semantically equivalent to the corresponding uniqueness-restricted existential quantifier “some unique author of *Waverley*,” in the sense of “something such that it, and nothing else, wrote *Waverley*.” The restricted quantifier falls under the purview of Russell’s general theory of descriptions. On Russell’s theory, then, “the author of *Waverley*” corresponds to the string “Someone is such that: he or she uniquely wrote *Waverley* and he or

she ... ,” making definite descriptions also “incomplete symbols” which have “no meaning in isolation.” The words “The author of *Waverley* is ingenious” are not directly about Walter Scott, but about the complex propositional function, *being a unique author of Waverley who is also ingenious*, expressing that this function yields a true proposition for at least one individual. There is nothing that the phrase “the author of *Waverley*” contributes on its own to this proposition.

As with “some S,” sentences that position a definite description within governing operators yield multiple readings. For example, “Jones believes the author of *Waverley* is not ingenious” may be analyzed alternatively, and nonequivalently, as: (i) “Someone uniquely wrote *Waverley* and Jones believes he is not ingenious”—that is, Jones believes of *Waverley*’s sole author that *he* is not ingenious (*wide scope*); (ii) “Jones believes: that someone both uniquely wrote *Waverley* and is not ingenious”—that is, Jones believes that whoever wrote *Waverley* single-handedly is not ingenious (*intermediate scope*); or (iii) “Jones believes: that no one both uniquely wrote *Waverley* and is ingenious” (*narrow scope*). The wide-scope reading is consistent with Jones’s belief not involving a conception of Scott as sole author of *Waverley*. The narrow-scope reading attributes a belief that is consistent with *Waverley* not having a sole author.

A definite description is said to be *proper* when there is someone or something that uniquely answers to the description, and is *improper* otherwise. Russell artificially, and misleadingly, extended Mill’s term “denotation” to the semantic relation that obtains between a proper definite description and the individual uniquely described, even though a definite description is supposed not to be a singular term. He might instead have called this relation “simulated denotation.” Russell retained the term “meaning” for semantic content.

Both the Millian theory and Russell’s theory deny that the individual that uniquely answers to a definite description is itself a component of the content of sentences involving the description. Those theories are able to solve the puzzles in the special case where the terms involved are definite descriptions rather than proper names, by reading sentences involving definite descriptions as containing propositions involving corresponding attributes or propositional functions. In particular, Russell’s claim that definite descriptions are not singular terms, but quantificational constructions, blocks substitutivity of equality, which is applicable only to singular terms, from licensing the substitution of “the first Postmaster General” for “the inventor of bifocals” in the

secondary-occurrence reading of “Jones believes that the inventor of bifocals was clever.” (By contrast, the envisioned substitution is indeed licensed by logical principles, including substitutivity as applied to variables, when the sentence takes on its primary-occurrence reading.)

Russell handled the same difficulties in the case of proper names (and such devices as demonstratives) through his thesis that names are ordinarily not used as “genuine names” (singular terms). Instead they were held to be “disguised” or “abbreviated” definite descriptions. The proposition expressed by a sentence involving a typical name is to be analyzed in accordance with Russell’s theory of descriptions. This blocks substitution in sentences such as “Jones believes that Hesperus appears in the evening.” Russell acknowledged the possibility of “names in the strict, logical sense” (logically proper names), which function in accordance with the naive theory. The class of admissible semantic contents for usable genuine names was severely limited by Russell’s principle of acquaintance, that every proposition one can grasp must be composed entirely of constituents with respect to which one has a special sort of intimate and direct epistemic access, (*direct acquaintance*). This restriction seems sufficient to prevent the puzzles from arising with logically proper names. (Russell did not countenance genuine names lacking a referent. Curiously, he claimed that singular existential and negative existential statements involving genuine names are without meaning. It would have been better to say that such sentences are always trivially true and trivially false, respectively.)

#### 4. FREGE’S THEORY OF SINN AND BEDEUTUNG

In his classic paper, “*Über Sinn und Bedeutung*” (1892), Frege abandoned the naive theory in favor of a richly elegant philosophy of semantics, which extends the Millian theory’s two-tiered semantics for definite descriptions and predicates to include all meaningful expressions. (Like Mill, and unlike Russell, Frege counted definite descriptions as singular terms.) Frege distinguished between the referent (*Bedeutung*) of an expression and its sense (*Sinn*). The sense of an expression contains a purely conceptual manner of presenting the name’s referent. Individuals that are not themselves senses—such as persons and even their sensations—cannot be constituents of a genuine Fregean sense. Furthermore, the sense of a singular term secures the term’s referent. An expression’s sense is a conception of something, and the expression’s referent, if there is one, is whatever uniquely fits the concept. The reference relation is thus the relative product of



a purely semantic relation (that between an expression and its sense) and a nonlinguistic relation (that between a sense and the object that fits it). Third, the sense of an expression is the semantic content. Expressions having the same sense must have the same referent, but importantly, expressions having the same referent may differ in sense. Frege illustrated his notion of sense by means of three lines that intersect in a single point. Then the phrases “the point of intersection of *a* and *b*,” “the point of intersection of *a* and *c*,” and “the point of intersection of *b* and *c*” converge in reference but diverge in sense.

The observation that proper names have a sense, as distinct from the referent, is tailor-made to solve both Frege’s Puzzle and the problem of how sentences involving nonreferring names can have content. Frege’s solution to the substitution problem is more complex. Crucial to Frege’s theory are the principles of extensionality and compositionality. They hold that the referent or sense, respectively, of a complex expression is a function of the referents or senses, respectively, of the component expressions. In the latter case Frege spoke metaphorically of the sense of a constituent expression as a *part* of the sense of the complex expression, so that the sense of the whole is composed of the senses of the parts.

Thus, if a constituent expression is replaced by one having the same sense, the sense of the whole is preserved, whereas if a constituent expression is replaced by one having the same referent but a different sense, the referent of the whole is preserved even though the sense is not. In particular, Frege held as a special case of extensionality that a compound expression having a nonreferring part must be nonreferring (“Sherlock Holmes’s older brother”). Frege argued, using extensionality, that the *cognitive value* (*Erkenntniswerte*) of a sentence is not the referent of the sentence, but is fixed by its sense, and that the referent of a sentence is one of two truth values, truth and falsity (“the true” and “the false”). Because a sentence refers to its truth-value, and a sentence involving a nonreferring name itself refers to nothing, such a sentence as “Sherlock Holmes is addicted to cocaine” is neither true nor false. (Frege held that the sentence *presupposes*, without asserting, that Sherlock Holmes exists.)

Frege argued that certain expressions create a special context in which subordinate expressions do not refer to their customary referent. When occurring within quotation marks (for example, in “direct discourse” reporting the words used by a speaker) an expression refers to itself. Analogously, expressions occurring subordinate to operators such as “Jones believes that” and “Jones said that” (the latter occurring in “indirect discourse” reporting the

content of a speaker’s utterance) refer to their *ungerade* (indirect, oblique) referent, which is the customary sense. Extensionality is to be understood as requiring the validity of substituting for a name in a sentence any expression having the same referent *in that same position*. (Scattered remarks suggest that Frege might have applied his doctrine of semantic shifting also to the problem of negative existentials.)

## 5. THE THEORY OF DIRECT REFERENCE

Despite a fundamental disagreement over the matter of singular propositions, there is common ground between Russell and Frege in regard to ordinary proper names. Both held a strong version of the theory that names are *descriptive*. On their view, if “St. Anne” is analyzable as “the mother of Mary,” it must be analyzable even further, because “Mary” is also supposed to be descriptive. But even “the mother of the mother of Jesus” must be in this sense further analyzable. If “ $\alpha$ ” is a nondescriptive singular term referring to Mary, then it may be said that the description “the mother of  $\alpha$ ” is *descriptive relative to Mary*. A *thoroughly descriptive* term is one that is descriptive but not descriptive relative to anything. The orthodox theory, shared by Russell and Frege, is the theory that proper names and similar devices are either thoroughly descriptive or descriptive relative only to items of direct acquaintance. Frege held the stronger thesis (which is retained by contemporary variants of Frege’s theory, such as that of John Searle) that proper names are thoroughly descriptive. Any departure from the stronger thesis would constitute a rejection of fundamental Fregean theory.

In recent philosophy the orthodox theory has been forcefully challenged, most notably by Keith Donnellan (1972), David Kaplan, Saul Kripke (1972, 1979), Ruth Barcan Marcus, and Hilary Putnam. These philosophers favor the theory of direct reference, which holds that proper names (and similar devices) are nondescriptive. Importantly, this theory does not deny that particular names may exhibit any or all of the three aspects of a Fregean sense mentioned in the previous section. What is denied is that the conceptual representation carried by a name secures the referent. But the direct-reference theory is significantly stronger than a simple denial of Russell’s doctrine that ordinary names are abbreviated definite descriptions. The theory holds that names are not even similar to definite descriptions. An immediate consequence is that a great many definite descriptions fail to be thoroughly descriptive or descriptive relative only to

items of direct acquaintance, because many contain names of ordinary individuals.

Three main kinds of arguments have been advanced in favor of the direct-reference theory. The modal and epistemological arguments are due chiefly to Kripke. Suppose for simplicity that the name “Shakespeare” simply means “the English playwright who wrote *Hamlet*, *Macbeth*, and *Romeo and Juliet*.” If the orthodox theory of names is correct, then the sentence, “Someone is Shakespeare iff he is an English playwright who is sole author of *Hamlet*, *Macbeth*, and *Romeo and Juliet*,” should express a necessary, a priori truth. On the contrary, however, it might have come to pass that Shakespeare elected to enter a profession in law instead of becoming a writer. Furthermore, it is possible, and is not ruled out solely by semantic reflection, that Francis Bacon should go on to write these plays. These intuitions are supported by a complementary intuition: that “Shakespeare” continues to refer to the same person even with respect to nonactual possible worlds in which Shakespeare lacks the distinguishing characteristics that people actually use to identify him—that is, even in discourse about such a counterfactual scenario. One important consequence of the direct-reference theory is that any proper name is a *rigid designator* (Kripke)—that is, it designates the same thing with respect to every possible world in which that thing exists and does not designate anything else with respect to other possible worlds.

One example of the semantic arguments for the direct-reference theory comes from Donnellan: According to the orthodox theory, the semantic content of the name “Thales” is determined by a description such as “the Greek philosopher who held that all is water.” But suppose that the man referred to by writers from whom the use of the name “Thales” derives never genuinely believed that all is water but was thought to, owing to some error or hoax, and that, by coincidence, there was a Greek hermit who did hold this bizarre view, though he bears no historical connection to anyone. Contrary to the orthodox theory, the name “Thales” would nevertheless refer to the first of the two. This argument seems to reveal also that the surrounding settings in which speakers find themselves, and not merely the concepts evoked in them, are crucial to determining the referents of the names they use. In a word, the securing of a referent for a name is a *contextual* phenomenon. Donnellan and Kripke have provided partial accounts of the securing of a referent for a name by means of historical chains of communication. Putnam has given a similar account of certain terms designating something by means of a “division of linguistic

labor.” Because of these accounts the direct-reference theory is sometimes called the causal theory of reference.

## 6. THE MILLIAN THEORY RECONSIDERED

What, then, is the semantic content of a name? It is tempting to answer that it is, or at least includes, a descriptive or conceptual “mode of presentation.” Although this proposal does not require that the associated mode of presentation secure the referent, it faces some of the same difficulties as the orthodox theory. A more general difficulty arises because the variations of the argument from Frege’s Puzzle against the naive theory and the Millian theory can be mounted against a wide variety of theories of semantic content, including Frege’s. The general strategy involved in that argument, however, seems to involve an error. This might be demonstrated through an application to a situation involving expressions for which it is uncontroversial that semantic content is exactly the same.

Suppose that foreign-born Sasha learns the words “ketchup” and “catsup” by actually consuming the condiment and reading the labels on the bottles. Suppose further that, because of his idiosyncratic experience, Sasha comes to believe that the substances so named are different condiments sharing a similar taste, color, and consistency. Whereas “Ketchup is ketchup” is uninformative for Sasha, “Catsup is ketchup” is informative. It would be a mistake, however, to conclude that “catsup” and “ketchup” differ in semantic content for Sasha. The terms are perfectly synonymous in English; indeed, they are arguably the same English word. Most English speakers learned one in a sort of ostensive definition, and the other as a strict synonym (or as an alternative spelling) of the first. If either may be learned by ostensive definition, then both may be—witness Sasha. This discredits the original argument from Frege’s puzzle.

One important consideration favoring the Millian theory over the orthodox theory comes by consideration of individual variables. Consider the following propositional-attitude attribution:

- (1) The planet Venus is an individual  $x$  such that Jones believes that  $x$  is a star.

It is characteristic of this *de re* (as opposed to *de dicto*) locution that it does not specify how Jones is supposed to conceive of Venus in believing it to be a star. The Orthodox Theorist contends that this is a result of the allegedly descriptive name “Venus” positioned outside of the scope of the nonextensional operator “Jones believes

that,” where it is open to substitution and to existential generalization. What is more significant, however, is that a nondescriptive singular term is positioned within the scope of the nonextensional context: the last occurrence of the variable “*x*” in (1). It follows by the principles of conventional semantics that (1) is true if and only if its component open sentence:

(2) Jones believes that *x* is a star

is true under the assignment of Venus as value for the variable. In turn, (2) is true under this assignment if and only if Jones believes the semantic content of the complete open sentence:

(3) *x* is a star

under the same assignment. But the fundamental characteristic of a variable with an assigned referent is that its semantic content is just its referent. This is precisely the point of using a variable rather than a definite description (such as “the first heavenly body visible at dusk”) within the scope of an attitude verb in a *de re* attribution. If a variable with an assigned value had, in addition to its value, a Fregean sense, then (3) would contain a specific general proposition, under the relevant assignment. If (1) is to fail to specify how Jones conceives of Venus, the content of (3) under the assignment of Venus to “*x*” can only be the singular proposition about Venus that it is a star. If the open sentence (3), under the assignment of Venus as the value of “*x*,” contains the singular proposition about Venus that it is a star, then so does the closed sentence “*a* is a star,” where “*a*” is an individual constant that refers to Venus. It is not the variability of a variable, but its structural simplicity, that gives it the feature that the variable’s semantic content, under an assignment of a referent, is just the assigned referent. (An exactly parallel argument proceeds using pronouns in place of variables, using “The planet Venus is such that Jones believes that *it* is a star.”)

It is important to note also that at least some aspects of the remaining puzzles would arise even in a language for which it was stipulated that the Millian theory is correct. Suppose, for example, that an authoritative linguistic committee that legislates the grammar and semantics of the language, and to which all speakers of the language give their cooperation and consent, decreed that proper names are to function exactly like the mathematician’s variables, “*x*,” “*y*,” and “*z*,” except that they are to remain constant. Ordinary speakers would presumably continue to regard co-referential names as not always interchangeable in propositional-attitude attributions. English speakers who use “ketchup” and “catsup” as exact synonyms may be inclined to assent to “Sasha believes that

ketchup is a sandwich condiment, but he does not believe that catsup is.” On philosophical reflection, however, it emerges that this expresses a logical impossibility. Similarly, speakers who agree to abide by the legislative committee’s decree about proper names might for independent pragmatic reasons be led to utter or to assent to such sentences as “Jones believes that Hesperus appears in the evening, but he does not believe that Phosphorus does.” Insofar as the same phenomena that give rise to the puzzles would arise even in the case of a language for which the Millian theory was true by fiat and unanimous consent (and do in fact arise with respect to such straightforward synonyms as “ketchup” and “catsup”), the puzzles cannot be taken as evidence against the Millian theory. A deeper understanding is needed of the puzzles, and a reexamination of the Millian theory in light of this deeper understanding.

*See also* Demonstratives; Indexicals; Quantifiers in Natural Language.

### Bibliography

- Church, A. *Introduction to Mathematical Logic*. Vol. 1, secs. 1–6. Princeton, NJ: Princeton University Press, 1956.
- Donnellan, K. “Proper Names and Identifying Descriptions.” In *Semantics of Natural Language*, edited by D. Davidson and G. Harman, 356–379. Dordrecht, Netherlands: D. Reidel, 1972.
- Frege, G. *Begriffsschrift*. 1879. Translated by G. Frege as “Conceptual Notation.” In his *Conceptual Notation and Related Articles*. Oxford: Oxford University Press, 1972.
- Frege, G. “Über Sinn und Bedeutung.” 1892. Translated as “On Sense and Meaning” or (preferably) “On Sense and Reference” or “On Sense and Nominatum.” Widely reprinted. Available in Harnish, *Basic Topics in the Philosophy of Language*. London: Prentice Hall/Harvester Wheatsheaf, 1993.
- Harnish, R. M. *Basic Topics in the Philosophy of Language*. London: Prentice Hall/Harvester Wheatsheaf, 1993.
- Kripke, S. *Naming and Necessity*. Cambridge, MA: Harvard University Press, 1972. Excerpted in Harnish, *Basic Topics in the Philosophy of Language*. London: Prentice Hall/Harvester Wheatsheaf, 1993.
- Kripke, S. “A Puzzle about Belief.” 1979. In Salmon and Soames, *Propositions and Attitudes*, 102–148. Also in Harnish, *Basic Topics in the Philosophy of Language*. London: Prentice Hall/Harvester Wheatsheaf, 1993.
- Mill, J. S. “Of Names.” In *A System of Logic*. 1893. Bk. 1, chap. 2. Excerpted in Harnish, *Basic Topics in the Philosophy of Language*. London: Prentice Hall/Harvester Wheatsheaf, 1993.
- Russell, B. “Knowledge by Acquaintance and Knowledge by Description.” In *Propositions and Attitudes*, edited by N. Salmon and S. Soames, 16–32. Oxford: Oxford University Press, 1988.

- Russell, B. "On Denoting." In *Mind*. 1905. Widely reprinted. Available in Harnish, *Basic Topics in the Philosophy of Language*. London: Prentice Hall/Harvester Wheatsheaf, 1993.
- Russell, B. *The Philosophy of Logical Atomism*, edited by D. Pears. La Salle, IL: Open Court, 1985.
- Salmon, N. "A Millian Heir Rejects the Wages of *Sinn*." In *Propositional Attitudes: The Role of Content in Logic, Language, and Mind*, edited by C. A. Anderson and J. Owens, 215–247. Stanford, CA: Center for the Study of Language and Information, 1990.
- Salmon, N. "A Problem in the Frege–Church Theory of Sense and Denotation." *Noûs* 27 (2) (June 1993): 158–166.
- Salmon, N. "Reference and Information Content: Names and Descriptions." In *Handbook of Philosophical Logic*. Vol. 4, *Topics in the Philosophy of Language*, edited by D. Gabbay and F. Guenther, 409–461. Dordrecht, Netherlands: D. Reidel, 1989.
- Salmon, N., and S. Soames, eds. *Propositions and Attitudes*. Oxford: Oxford University Press, 1988.
- Searle, J. "Proper Names." *Mind* 67 (266) (1958): 166–173.

*Nathan Salmon (2005)*

## PROPERTIES

Our every assertion or thought involves "properties" or relations. Most simply, we predicate some property of some thing: Earth is round. Sometimes we refer to properties by name or by description: Red is the color of blood. Sometimes our quantifiers range over properties: Galaxies come in many shapes and sizes.

This familiarity with properties, however, does not reveal what properties are. Indeed, the question is equivocal, both in ordinary and in philosophical discourse. There are different conceptions of properties, equally legitimate, corresponding to the different roles that properties have been called upon to play (Bealer 1982; Lewis 1983, 1986). And for each conception there are different theories as to what sort of entity, if any, is best suited to play the role. The most fundamental division is between abundant and sparse conceptions of properties. On an abundant conception every meaningful predicate expresses some property or relation, including "is blue or round," "is on top of a turtle," "is identical with the planet Mars"; a property's instances need not resemble one another in any intrinsic respect. Abundant properties are needed to serve as "meanings," or components of "meanings," in a compositional semantics for language. On a sparse conception of properties a predicate expresses a property only if the objects satisfying the predicate resemble one another in some specific intrinsic respect; perhaps "has unit positive charge" and "is ten kilograms

in mass" are examples. Sparse properties are needed to provide an objective basis for the scientist's project of discovering the fundamental classifications of things and the laws that govern them. Properties, whether abundantly or sparsely conceived, are neither language- nor mind-dependent: They existed before there were beings to talk and think about them; they would have existed even had there never been such beings.

In this entry only conceptions of properties are explicitly distinguished and discussed, although much of what is said applies also to relations and to propositions. Other philosophers' terms for *property* in the abundant sense include *attribute* (Quine 1970), *propositional function* (Russell 1919), and *concept* (Bealer 1982, Frege 1884); *universal* and *quality* have for the most part been interpreted sparsely. Ordinary language allows abundant or sparse readings of *characteristic*, *feature*, *trait*, and more.

## ABUNDANT CONCEPTIONS OF PROPERTIES

How abundant are the properties on the abundant conception? Whenever there are some things, no matter how scattered or dissimilar from one another, there is the abundant property of being one of those things. Thus, for any class of things, there is at least one abundant property had by all and only the members of that class. It follows that there are at least as many abundant properties as classes of things and that the abundant properties outrun the predicates of any ordinary language. (There are non-denumerably many classes of things—assuming an infinity of things—but at most denumerably many predicates in any ordinary language.) Abundant properties, owing to their very abundance, must be transcendent, rather than immanent: They are not present in their instances as constituents or parts. It is not plausible to suppose that an object has a distinct constituent for each and every class to which it belongs.

If we say that whenever there are some things, there is exactly one property had by all and only those things, then a property may be identified with the class of its instances. For example, the property of being human may be identified with the class of human beings. But there is a well-known objection to this identification (Quine 1970). Consider the property expressed by "is a creature with a heart" and the property expressed by "is a creature with kidneys." If properties are "meanings," or semantic values, of predicates, then the properties expressed by these two predicates are distinct. Yet, these predicates, we may suppose, are coextensive: As a matter of fact, any creature with a heart has kidneys, and vice versa; the class

of creatures with a heart is identical with the class of creatures with kidneys. Thus, distinct properties correspond to the same class and cannot be identified with that class.

Different responses to the objection invoke different criteria of individuation for properties, that is, different criteria for deciding when properties, introduced, say, via predicates that express them, are one and the same. One response simply denies that “is a creature with a heart” and “is a creature with kidneys” express distinct properties. More generally, properties expressed by coextensive predicates are identical. Call this an extensional conception of properties. A property so conceived may be identified with the class of its instances. Extensional conceptions of properties are adequate to the semantic analysis of mathematical language and extensional languages generally (Tarski 1946).

A second response holds that “is a creature with a heart” and “is a creature with kidneys” express distinct properties, because it is logically possible for something to satisfy one predicate without satisfying the other. On this response properties expressed by necessarily coextensive predicates are identical; properties expressed by accidentally coextensive predicates are distinct. Call this an intensional conception of properties. If one accepts the standard analyses of logical possibility and necessity in terms of *possibilia*, then a property, on the intensional conception, may be identified with the function that assigns to each possible world the set of possible objects that has the property at the world. If one holds that each object exists at, and has properties at, only one world, then a property may more simply be identified with the class of (actual and) possible objects that has the property (Lewis 1986). Properties, on the intensional conception, are appropriate semantic values for predicates of (standard) modal languages and intensional languages generally (Carnap 1947, Kripke 1963).

A third response holds that the properties expressed by “is a creature with a heart” and “is a creature with kidneys” are distinct because they are structured entities with different constituents: The property expressed by “is a creature with a heart” has the property expressed by “is a heart” as a constituent; the property expressed by “is a creature with kidneys” does not. On this response properties have a quasi-syntactic structure that parallels the structure of predicates that express them. Call two predicates isomorphic if they have the same syntactic structure and corresponding syntactic components are assigned the same semantic values. On a structured conception of properties, properties expressed by isomorphic predicates are identical; properties expressed by nonisomorphic

predicates are distinct. (Structured conceptions are sometimes called hyperintensional because they allow necessarily coextensive predicates to express distinct properties.) Structured conceptions subdivide according to whether the unstructured semantic values are intensional or extensional and according to whether the relevant structure is surface grammatical structure, or some hypothetical deep structure, or structure after analysis in terms of some chosen primitive vocabulary. Structured properties may be identified with sequences of unstructured properties and other unstructured semantic values. Structured properties, on one version or another, have a role to play in the semantic analysis of propositional attitudes and of hyperintensional languages generally (Carnap 1947, Cresswell 1985).

Thus far, this entry has assumed that predicates of ordinary language are satisfied by objects once and for all. In fact, most ordinary language predicates are tensed; they may be satisfied by objects at some times but not at others. For example, “is sitting” is true of me now, but was false of me ten minutes ago. On a tensed conception of properties, whether or not a property holds of an object may also be relative to times. Most simply, tensed properties may be identified with functions from times to untensed properties. Tensed properties may be taken as semantic values for tensed predicates.

We have, then, a plurality of abundant conceptions of properties. Which is correct? One need not and should not choose. A plurality of conceptions is needed to account for the multiple ambiguity in our ordinary talk of properties. And it seems that both structured and intensional conceptions are needed for compositional semantics: Structured properties are needed to provide distinct semantic values for predicates, such as “is a polygon with three sides” and “is a polygon with three angles,” that are necessarily coextensive without being synonymous; intensional properties are needed to provide distinct semantic values for unstructured predicates that are accidentally coextensive. To accept a plurality of conceptions, it suffices to find, for each conception, entities that satisfy that conception’s criteria of individuation.

Realists with respect to some conception of properties hold that entities satisfying the individuation criteria for the conception exist. Realists divide into reductionists and antireductionists. Reductionists identify properties, under the various conceptions, with various set-theoretic constructions (in ways already noted): class, functions, or sequences of actual or possible objects (Lewis 1986). Antireductionists reject some or all of these identifications. For some antireductionists, classes are suspect or

esoteric entities; classes are to be explained, if at all, in terms of properties, not vice versa (Bealer 1982, Russell 1919). For other antireductionists the problem is not with classes, but with the possibilia that comprise them (on intensional conceptions). Possible but nonactual entities are to be explained, if at all, in terms of uninstantiated properties, not vice versa (Plantinga 1976). According to the antireductionist, properties are basic or primitive; it is merely posited that there are entities satisfying the appropriate individuation criteria. Some entities, after all, must be taken as basic; according to the antireductionist, properties are an acceptable choice.

Eliminativists hold that, strictly speaking, there are no properties. They take aim, typically, at intensional conceptions, at conceptions with modal criteria of individuation. They claim that modal notions, such as logical possibility and necessity (whether taken as primitive or analyzed in terms of possibilia), incorrigibly lack the clarity and precision required of a rigorous scientific semantics or philosophy (Quine 1970). Eliminativists have the burden of showing how ordinary and philosophical discourse ostensibly referring to properties can be paraphrased so as to avoid such reference; or, failing that, of showing that such discourse is dispensable, merely a *façon de parler*.

### SPARSE CONCEPTIONS OF PROPERTIES

On an abundant conception any two objects share infinitely many properties and fail to share infinitely many others, whether the objects are utterly dissimilar or exact duplicates. On a sparse conception the sharing of properties always makes for genuine similarity; exact duplicates have all of their properties in common. Whatever the sparse properties turn out to be, there must be enough of them (together with sparse relations) to provide the basis for a complete qualitative description of the world, including its laws and causal features. The sparse properties correspond one-to-one with a select minority of the abundant properties, on some intensional conception. (“Intensional,” because distinct sparse properties may accidentally be instantiated by the same objects.) Those abundant properties that correspond to sparse properties are called natural (or perfectly natural, since naturalness presumably comes in degrees; Lewis 1983, 1986). The naturalness of properties is determined not by our psychological makeup, or our conventions, but by nature itself.

How sparse are the properties, on a sparse conception? First, there is the question of uninstantiated properties. If sparse properties are transcendent, there is no

difficulty making room for uninstantiated sparse properties; perhaps uninstantiated sparse properties are needed to ground laws that come into play only if certain contingent conditions are satisfied (Tooley 1987). If, on the other hand, sparse properties are immanent, are present in their instances, then uninstantiated sparse properties must be rejected, because they have nowhere to be (Armstrong 1978, 1989). Of course, uninstantiated sparse properties may nonetheless possibly exist, where this is understood according to one’s favored interpretation of modality.

Second, there is the question of the compounding of sparse properties (and relations). Disjunctions and negations of natural properties are not themselves natural: Their instances need not resemble one another in any intrinsic respect. For example, instances of the property having-unit-positive-charge-or-being-ten-kilograms-in-mass need not resemble one another in either their charge or their mass. It follows that there are no disjunctive or negative sparse properties (Armstrong 1978).

The case of conjunctive sparse properties is less clear. There are two views. According to the first, since instances of a conjunction of natural properties, such as having-unit-positive-charge-and-being-ten-kilograms-in-mass, resemble one another in some—indeed, at least two—intrinsic respects, there exists a sparse property corresponding to the conjunction. According to the second view, the sparse properties must be nonredundant; they must be not only sufficient for describing the world but minimally sufficient. On this view conjunctive sparse properties are excluded on grounds of redundancy: A putative conjunctive sparse property would hold of an object just in case both conjuncts hold.

Similarly, structural sparse properties, such as being-a-molecule-of-H<sub>2</sub>O, may be admitted on the grounds that they make for similarity among their instances. Or they may be excluded on grounds of redundancy: A putative structural sparse property would hold of an object just in case certain other sparse properties and relations hold among the object and its parts. But the exclusion of structural (and conjunctive) sparse properties faces a problem. It rules out a priori the possibility that some properties are irresolvably infinitely complex: They are structures of structures of structures, and so on, without ever reaching simple, fundamental properties or relations (Armstrong 1978). A sparse conception that allowed for this possibility would have to allow some redundancy; and if some redundancy, why not more? This suggests that conjunctive and structural sparse properties should generally be admitted. (An alternative treatment makes use of degrees

of naturalness and has it that conjunctive and structural properties are natural to some lesser degree than the properties in terms of which they are defined; a world with endless structure has no perfectly natural properties.)

If structural sparse properties are admitted, the sparse properties will not be confined to fundamental physical properties; there will be sparse properties of macroscopic, as well as microscopic, objects. For example, the sparse properties will include specific shape-and-size properties, such as being-a-sphere-ten-meters-in-diameter (which are arguably structural properties definable in terms of sparse distance relations). However, the vast majority of ordinary-language predicates—“is red,” “is human,” “is a chair,” to name a few—fail to express natural properties to which sparse properties correspond; rather, these predicates express properties that, when analyzed in fundamental physical terms, are disjunctive (perhaps infinitely so) and probably extrinsic. (This judgment could be overturned, however, if there are irreducible natural properties applying to macroscopic objects—most notably, irreducible phenomenological properties of color, sound, and such.)

What are the properties on a sparse conception? There are three principal theories (or clusters of theories, since they each subdivide). According to the first, the properties sparsely conceived are just some of the properties abundantly conceived: The properties that are perfectly natural. What makes some properties natural and others unnatural? One version of the theory simply takes naturalness to be a primitive, unanalyzable distinction among abundant properties (Quinton 1957; see also Armstrong 1989, Lewis 1986). But since a property is natural in virtue of the resemblances among its instances, it might seem more appropriate to take instead some relation of partial resemblance as primitive and to define naturalness in terms of resemblance. The resulting version, called resemblance nominalism, can be worked out in different ways with different primitive resemblance relation (Price 1953; see also Armstrong 1989, Goodman 1951; Lewis 1983). The chief objection to the view is that partial resemblance between ordinary objects, no less than naturalness of properties, cries out for analysis. When two objects partially resemble one another, the objection goes, they must have constituents that exactly resemble one another, perhaps constituents that are literally identical. More generally, it is argued, properties must be constituents of objects if properties are to play a role in the explanation of the natures and causal powers of objects; one cannot explain an object’s nature or causal powers by

invoking a class to which it belongs. Sparse properties, then, must be immanent, not transcendent, entities.

What are these constituents of ordinary things? Not ordinary spatial or temporal constituents—or, at least, not always. For even an object with no spatial or temporal extension might have a complex nature and stand in relations of partial resemblance. If sparse properties are immanent, then they must be nonspatiotemporal constituents of things. There are two prominent theories as to the nature of these constituents. The first theory takes them to be universals (Armstrong 1978, 1989.) They are repeatable: Each of them is, or could be, multiply instantiated. And they are wholly present in their instances: An immanent universal is located—all of it—wherever each of its instances is located. When objects resemble one another by having a sparse property in common, there is something literally identical between the objects. It follows that universals fail to obey commonsense principles of location, such as that nothing can be (wholly) in two places at the same time. But that is no objection. Such principles were framed with particulars in mind; it would beg the question against universals to require them to meet standards set for particulars.

On the other theory of sparse properties as immanent, the nonspatiotemporal constituents of ordinary particulars are themselves particulars, called tropes (Armstrong 1989, Lewis 1986, Williams 1966) or abstract particulars (Campbell 1981). When ordinary particulars partially resemble one another by having some sparse property—say, their mass—in common, then there are distinct, exactly resembling, mass tropes as constituents of each. On a trope theory sparse properties can be identified with maximal classes of exactly resembling tropes (perhaps including merely possible tropes). Exact resemblance between tropes is taken as primitive by trope theory; but it is a simple and natural primitive compared to the partial resemblance relation taken as primitive by an adequate resemblance nominalism.

A possible disadvantage of a universals theory is that it requires two fundamentally distinct kinds of entities: universal and particulars. An ordinary particular cannot simply be identified with a bundle of coinstantiated universals, lest numerically distinct but qualitatively identical particulars be identified with one another. On a universals theory there must be some nonqualitative, nonrepeatable constituent of ordinary particulars to ground their numerical identity. A trope theory, on the other hand, needs only tropes to make a world. Ordinary particulars can be identified with bundles of coinstantiated tropes; numerically distinct but qualitatively identical

particulars are then bundles of numerically distinct but exactly resembling tropes.

The great advantage of a universals theory is that it promises to analyze all resemblance in terms of identity: Exact resemblance is identity of all qualitative constituents; partial resemblance is partial identity, identity of at least one qualitative constituent. But it is unclear whether the promise can be kept. Objects instantiating different determinates of a determinable—such as unit-positive and unit-negative charge—seem to partially resemble one another by both being charged without there being any analysis of this resemblance in terms of the identity of constituent universals or, for that matter, the exact resemblance of constituent tropes. A universals theory and a trope theory would then have to fall back upon primitive partial resemblance between universals, or tropes. Some of the advantages of these theories over resemblance nominalism would be forfeited.

Of the three basic theories of sparse properties—resemblance nominalism, a theory of immanent universals, and a theory of tropes—only one can be true; the theories posit incompatible constituent structure to the world. However, assuming each theory is internally coherent, and adequate to the needs of science, the question arises, What sort of evidence could decide between them? It seems that a choice between the theories will have to be made, if at all, on the basis of pragmatic criteria such as simplicity, economy, and explanatory power. There is as yet no philosophical consensus as to what that choice should be.

**See also** Armstrong, David M.; Carnap, Rudolf; Eliminative Materialism, Eliminativism; Frege, Gottlob; Goodman, Nelson; Kripke, Saul; Metaphysics; Quine, Willard Van Orman; Realism; Reduction; Reductionism in the Philosophy of Mind; Russell, Bertrand Arthur William; Tarski, Alfred.

### **Bibliography**

- Armstrong, D. M. *Universals and Scientific Realism*. 2 vols. Cambridge, U.K.: Cambridge University Press, 1978.
- Armstrong, D. M. *Universals: An Opinionated Introduction*. Boulder, CO: Westview Press, 1989.
- Bealer, G. *Quality and Concept*. Oxford: Clarendon Press, 1982.
- Campbell, K. "The Metaphysics of Abstract Particulars." In *Midwest Studies in Philosophy*. Vol. 6. Minneapolis: University of Minnesota Press, 1981.
- Carnap, R. *Meaning and Necessity*. Chicago: University of Chicago Press, 1947.
- Cresswell, M. *Structured Meanings*. Cambridge, MA: MIT Press, 1985.

- Frege, G. *Die Grundlagen der Arithmetik* (1884). Translated by J. L. Austin as *The Foundations of Arithmetic*. Oxford: Blackwell, 1950.
- Goodman, N. *The Structure of Appearance*. Cambridge, MA: Harvard University Press, 1951.
- Kripke, S. "Semantical Considerations on Modal Logic." *Acta Philosophica Fennica* 16 (1963): 83–94.
- Lewis, D. "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61 (1983): 343–377.
- Lewis, D. *On the Plurality of Worlds*. Oxford: Blackwell, 1986.
- Plantinga, A. "Actualism and Possible Worlds." *Theoria* 42 (1976): 139–160.
- Price, H. H. *Thinking and Experience*. London: Hutchinson's University Library, 1953.
- Quine, W. V. O. *Philosophy of Logic*. Englewood Cliffs, NJ: Prentice Hall, 1970.
- Quinton, A. "Properties and Classes." *Proceedings of the Aristotelian Society* 58 (1957): 33–58.
- Russell, B. *Introduction to Mathematical Philosophy*. London: Allen and Unwin, 1919.
- Tarski, A. *Introduction to Logic*. 2nd ed. Oxford, 1946.
- Tooley, M. *Causation*. Oxford: Clarendon Press, 1987.
- Williams, D. C. "The Elements of Being." In *The Principles of Empirical Realism*. Springfield, IL: Thomas, 1966.

**Phillip Bricker (1996)**

## PROPERTY

The institution of property has interested social philosophers in part, at least, because it raises issues of justice. Like government, it is practically universal but varies enough in its particular arrangements to suggest the question What criteria are relevant in assessing the relative merits of various arrangements? Again, because it discriminates between rights and fortune, it invites moral criticism and the demand for justification.

Many of the classical accounts of the origin and function of private property have taken for granted that in nature all things were held "in common." This phrase, however, is ambiguous, for it often meant not a system regulating the use of goods by general agreement but a condition where, there being no rules, everything was *res nullius* (a thing belonging to no one) and the concept "property" was consequently irrelevant. How, then, it was asked, would humans come to appropriate the land and its fruits? How could such appropriation be justified? What would be rational grounds for claiming exclusive possession? And could there be any limit on people's right to do what they would with their own?



## THEORIES OF PROPERTY

According to the Church Fathers, property was both the consequence and the social remedy for the sin of covetousness that came with the Fall. But since owners have appropriated what at one time belonged to all, they have a duty to administer it for the benefit of all. “Our property,” said Gregory the Great, “is ours to distribute, but not ours to keep.” The concept of the owner as steward is the core of the traditional Christian view of property.

**NATURAL LAW AND CONVENTIONALISM.** By the seventeenth century, property rights came to be grounded in the needs and accomplishments of the individual owner, and ownership implied a natural right to enjoy and dispose of its objects, limited only by the duty to respect the rather narrowly defined interests of others. In John Locke’s account, property as an institution is explained by human needs. Although God gave the earth and all its fruits to all people to preserve their lives, still this meant one’s making the fruits of the earth exclusively his own, if only by eating them. However, what in nature entitled one to call something one’s own was that one made the effort to make it so. To add one’s labor to a *res nullius* was to create a title to the whole product. Locke limited this title to whatever one could use before it spoiled; appropriation for waste would be illegitimate.

To appropriate an object implied for Locke not merely a right to enjoy it but also to alienate it at will, so that although the appropriation of *res nullius* could legitimately be effected only by labor, the title, once established, could be freely transferred. It is questionable, however, whether Locke was justified in assuming that because we may appropriate what we need from the common stock, we may therefore transfer what we acquire, but do not need, to whomsoever we choose. Locke needed this right, however, if his theory was not to suggest, as did certain later writers on economic justice, that the laborer was entitled to the entire fruits of his master’s fields, if not to the fields themselves. For where all land had long been appropriated, the titles of present owners would depend entirely on the legitimacy of such transfers in the past. So, since the land was no longer *res nullius*, all the laborer could claim was the value of his labor in wages. Moreover, in a market economy and with the introduction of money, wealth might be accumulated and stored indefinitely without spoiling; furthermore, since money had only a conventional value, hoarding it deprived no one of anything of natural value, and its distribution must be taken to be by common consent. Having accounted, then, for the existence of property, and for

existing titles, with a theory of natural right, Locke overlaid the theory with a conventionalist theory that neutralized the limitations on appropriation that the original theory prescribed.

Nature and convention are to be found similarly blended, if in varying proportions, in Hugo Grotius, Samuel von Pufendorf, and William Blackstone. In Immanuel Kant, too, there is a blend. Kant deduced the principle of first occupier from the autonomy of the will but conceded that only a universal legislative will—the civil state—could give binding force to the intention to appropriate.

**UTILITARIAN POSITIONS.** According to David Hume, a man’s creation ought to be secured to him in order to encourage “useful habits and accomplishments.” Inheritance and the right to alienate were alike valuable as incentives to or conditions for useful industry and commerce. Property rested on convention in the sense of rules upheld by common interests commonly perceived. It was a law of nature, too, but in the sense that men were sufficiently alike the world over for the same general arrangements to be equally to the public advantage. Hume’s argument, then, also blends natural law doctrine with conventionalism but reduces both to utilitarianism.

Jeremy Bentham did little more than elaborate Hume’s arguments. However, by introducing considerations of utility, Hume and Bentham pointed the way for criticism of the distribution of private property and, indeed, of the institution itself. Already in 1793 William Godwin was arguing that in a consistent application of the principle of utility “every man has a right to that, the exclusive possession of which being awarded to him, a greater sum of benefit or pleasure will result, than could have arisen from its being otherwise appropriated” (*Political Justice*, Book 8). J. S. Mill, though broadly committed to a belief in private property, held that, in the case of land at any rate, private ownership must be conditional on its expediency; the rights associated with it, especially the right to exclusive access and enjoyment, ought to be limited to whatever was required to exploit it efficiently. Mill recognized that the rights of property were not an inseparable bundle, to be justified en bloc; each constituent right had to be independently justified on grounds of utility.

However, Mill’s belief that the institution of property would be justified provided that it guaranteed to individuals the fruits of their labor and abstinence is open to question. In a complex industrial society, “the fruits of one’s labor” can mean only the value of a given worker’s

contribution to the finished product. But value derives from the relations of supply and demand, both for the commodity and for labor of the various kinds needed to produce it. “The fruits of one’s labor,” understood as one’s share in a social dividend, will depend not only on one’s efforts but also on the number of other people available to do the same job and on how badly consumers want it done. If for the time being a particular skill is in short supply, is it self-evident that this increases the value of its fruits or that those who have it should be the better off for it?

**MARXIST AND HEGELIAN CRITIQUES.** Again, the exclusive claims of labor take no account of what men owe to others and to the social interest. Émile Durkheim, for instance, objected that “it is not enough to invoke the rights that man has over himself: these rights are not absolute but limited by the claims of the moral aims, in which a man has to cooperate.” Karl Marx was equally critical of the German Social Democrats’ Gotha program of 1875, which claimed that labor should receive its produce “unabridged and in equal right.” He charged that this formula ignored the need for capital replacement and development, social services, and the support of the incapable. In any case, he said, distribution proportional to contribution would still be only partial justice, bearing in mind differences in natural capacity on the one hand and need on the other. In the truly cooperative society, based on common ownership of the means of production, individual labor would be impossible to separate out, and distribution would be according to need alone. This would be possible, however, only because labor would have ceased to be a burden and would have become “life’s principal end.”

This last condition suggests why, in a period when hedonistic premises underlay a great deal of psychology, ethics, and economics, the necessary relation between labor and property should have been so generally accepted. On the assumption that work was painful, the only conceivable reason for working was a greater pleasure expected from its fruits. Marx argued that this account of labor was neither an explanation nor a justification, but a consequence, of the system of private property. The worker was alienated from his work, which appeared to him not as a fulfillment but as a burden; he was alienated, too, from the product of his work, which, passing to his employer in surplus value, confronted him as capital—that is, as an instrument of his own bondage.

Despite the stress on labor as the source of value that Marx shared with the English utilitarians and econo-

mists, his account of property derives at least as much from G. W. F. Hegel as from the English school. Like Kant, Hegel regarded property as necessary not because it helped to satisfy human needs but because “a person must translate his freedom into an external sphere in order that he may achieve his ideal existence” (*Philosophy of Right*, Sec. 41); because “property is the first embodiment of freedom and so is in itself a substantive end” (Sec. 45). Plato erred, in Hegel’s view, in denying private property to the guardians, for he was denying them the conditions necessary for giving concrete realization to their personalities and wills.

Marx and Hegel are alike in seeing the human will objectifying itself in its acquisitions and creations. If for Marx the process is not rationalizing and liberating but alienating and enslaving, it is because the property created is not and cannot be the worker’s own. The laborer can transcend this alienation only in the communist society, in which, like Plato’s guardians or the members of a monastic community, he gets caught up in a common enterprise where “mine” and “thine” are of no account because life is more than the satisfaction of material needs. In a world in which “sharing in” counted for more than “sharing out,” property—like justice—would present no problems.

**ECONOMIC AND SOCIAL SIGNIFICANCE.** In the course of the past century, legal and social philosophers (Léon Duguit and Karl Renner, for example) have come to think of property increasingly as an institution with social functions and not, like Locke, as simply a guarantee of individual interest. Moreover, because property entails inequalities in power, in claims on the social product, in social status, and in prestige, it must be justified, and not merely in terms of the interests or natural rights of its immediate beneficiaries.

It is difficult, however, to see how any one theory could apply generally to all forms of private property and include all rights of ownership. Individual control of productive resources raises very different issues from the exclusive right to enjoy consumer goods such as clothes and furniture. The right to control the use of mines and factories is not really an instance of the right of a Kantian rational and autonomous being to manipulate mere things for his own needs; it is also an exercise of power over other people. According to A. A. Berle, the United States is gradually extending to such property the limitations traditionally applied to state action in order to protect individual freedom.

Again, could one justify one's title to dividends on the ground that instead of enjoying the fruits of one's labor one had invested them? And would the same justification extend to a corporate title to the yield on investments financed out of undistributed profits? Such claims have certainly flourished under the umbrella of natural rights; but it is difficult to see how any but a utilitarian argument could seriously be proposed in defense of such arrangements.

## ANALYSIS OF OWNERSHIP

Talk of property often seems to be talk about things. Things constitute property, however, only inasmuch as they can be assigned to owners; to own something is to have, in respect to it, certain rights and liabilities vis-à-vis other persons or the public at large. Ownership, therefore, is a normative relation or a complex of such relations between owner, object, and third parties, and to refer to something as "property" is to locate it as a term in such a relationship. Some jurists, indeed, insist that *property* refers not to things at all but, rather, to a bundle of rights. And this is obviously true of income titles, such as securities and annuities, and of rights of control over "intellectual property," such as patent rights and copyrights; these are "things" only in a very abstract sense, as characteristic complexes of normative relations.

As the objects of property are diverse, so also are the rights constituting it. Landowners' rights are necessarily different from copyright owners', and the owner of a gun does not have the same unrestricted use and control of it as the owner of a table has of the table. Jurists have nevertheless tried to identify some right necessary to ownership. The rights of exclusive use, possession, or alienation seem to be likely candidates, but each can conceivably be detached (for example, by a lease or an easement, under the terms of a trust, or, in former times, by entail) without the owners' losing property in the object. Accordingly, Sir Frederick Pollock suggested that "we must look for the person having the residue of all such powers when we have accounted for every detached and limited portion of it." But this residue, as held, say, by a ground landlord with a thousand-year tenant, may be very slender indeed, and the owner to whom all the detached rights will revert when the encumbrances reach the end of their term will certainly not be the present owner.

A. M. Honoré suggests a way out of these difficulties by concentrating not on the difficult exceptions but on the standard instance. He defines ownership as "those legal rights, duties and other incidents which apply, in the ordinary case, to the person who has the greatest interest

in a thing admitted by a mature legal system." Among the characteristic features are the right to possess and to be secure in possession, to use and to manage the property, to enjoy income arising from it and to alienate, consume, waste, or destroy the capital, and to transmit ownership to one's successors indefinitely; the absence of a fixed date on which the owners' interests terminate; the prohibition of harmful use; the liability of the property to execution for debt or insolvency; and the reversion to the owner on the termination of whatever lesser interests (leases, usufructs) encumber the property. Now, to say that *A* is the owner of *x* is not necessarily to say that he is the present subject of all these incidents; however, provided the kind of property in question can intelligibly be said to be the object of them and in the absence of special conditions or reservations, it is reasonable to infer that he is.

The Scandinavian legal realists—Karl Olivecrona and Alf Ross, for example—have been more radical in their analyses. According to Ross, ownership is "solely ... a tool in presentation." Theoretically, one could enunciate a mass of directives to judges, each consisting of a conditioning fact or facts (*F*) and a legal consequence (*C*), such as (1) if a person has lawfully purchased a thing ( $F_1$ ), judgment for recovery of possession should be given in his favor ( $C_1$ ); (2) if a person by prescription has acquired a thing and raised a loan that is unpaid ( $F_2$ ), the creditor should be given judgment for satisfaction out of the thing ( $C_2$ ); and so on. Now, to introduce "ownership" is not, according to Ross, to add something that *accounts* for the connections between the *F*'s and the *C*'s but merely to indicate the systematic connection between them such that  $F_1, F_2, F_3, \dots, F_p$  severally and collectively entail the totality of legal consequences  $C_1, C_2, C_3, \dots, C_n$ . The word *ownership* in Ross's view is "without any semantic reference whatever"; it serves only to reduce the complexity of particular rules to a systematic order. There is nothing beyond or in addition to the rules.

Now, it is certainly true that only confusion can result from trying to identify some special kind of a thing, or some special quality of things, which is called "property." Nevertheless, "ownership" does not always imply the same bundle of rights. The possible conditioning facts and the legal consequences are not the same for every case in which one may say that *X* is the owner of *P*. And, therefore, since the relevant rules do not have the rigorous relation to one another that Ross suggests, one can identify them as the rules of property (as distinct from, say, personal rights) only by recognizing some sort of family resemblances between them. Indeed, the terms Ross uses in exemplifying his conditioning facts—"purchase,"

“occupation of *res nullius*,” “acquisition by prescription”—are obviously already impregnated with ownership; to purchase something, for example, is to give money for it—that is, on the understanding that one acquires not merely possession, but also owners’ rights, over it.

Deciding who is the owner of a piece of property is, of course, to decide on the basis of certain facts where certain powers and liabilities lie. But to reduce a legal concept like property to a finite set of directives to judges ignores the fact that judges are constantly having to reshape the rules in the very process of applying them. If the rules of ownership are treated as a more or less arbitrary agglomeration, it is difficult to see how judges could make rational decisions at all.

Ross’s bundle of conditioning facts and legal consequences is significant, however, because it suggests how one goes about constructing a paradigm case of ownership, or, rather, a family of paradigms related by the fact that different conditioning facts entail broadly similar legal consequences. Deciding ownership in an atypical case would then involve deciding whether it can be assimilated to any of the available paradigms even though some characteristic ownership features are absent or other features that are out of character are present. A judge may have all kinds of reasons for making or refusing such an assimilation; but it is difficult to see how the problem could be presented to him at all without presupposing the standard cases of ownership as agreed starting points for discussion.

*See also* Bentham, Jeremy; Durkheim, Émile; Godwin, William; Grotius, Hugo; Hegel, Georg Wilhelm Friedrich; Hume, David; Justice; Kant, Immanuel; Locke, John; Marx, Karl; Mill, John Stuart; Natural Law; Patristic Philosophy; Plato; Pufendorf, Samuel von; Rights; Utilitarianism.

### **Bibliography**

Dias, R. W. M. *A Bibliography of Jurisprudence*. London, 1964.  
Extensive annotated bibliography, mainly from the juristic standpoint.

#### GENERAL HISTORIES

Bartlett, J. V. et al., eds. *Property, Its Duties and Rights*. London: Macmillan, 1915.

Grace, Frank. *The Concept of Property in Modern Christian Thought*. Urbana: University of Illinois Press, 1953.

Macpherson, C. B. *The Political Theory of Possessive Individualism: Hobbes to Locke*. Oxford: Clarendon Press, 1962.

Schlatter, Richard. *Private Property: The History of an Idea*. London: Allen and Unwin, 1951.

#### CLASSICAL THEORIES

Hegel, G. W. F. *Grundlinien der Philosophie des Rechts*. Berlin, 1821. Translated, with notes, by T. M. Knox as *Hegel’s Philosophy of Right*. Oxford: Clarendon Press, 1942.

Hume, David. *Treatise of Human Nature* (1640), edited by L. A. Selby-Bigge. 2nd ed. Oxford: Clarendon Press, 1896.

Locke, John. *Two Treatises of Government* (1690). Edited by Peter Laslett. Cambridge, U.K.: Cambridge University Press, 1960.

Marx, Karl. *Critique of the Gotha Programme*. In *Karl Marx and Friedrich Engels: Selected Works*. 2 vols. Moscow, 1949–1951. Vol. II.

Marx, Karl. “Oekonomisch-philosophische Manuskripte aus dem Jahre 1844.” In *Marx-Engels Gesamtausgabe*, edited by D. Riazanovsky and A. Adoratski. Berlin, 1932. Part I, Vol. III, edited and translated by T. B. Bottomore as “Economic and Philosophical Manuscripts” in *Karl Marx: Early Writings*. London, 1963.

Mill, J. S. *Principles of Political Economy*. 7th ed. London, 1871.

#### ECONOMIC AND SOCIAL SIGNIFICANCE

Berle, A. A. “Property, Production and Revolution.” *Columbia Law Review* 65 (1) (1965): 1–20.

Berle, A. A., and G. C. Means. *The Modern Corporation and Private Property*. New York: Macmillan, 1933.

Duguit, Léon. *Les transformations générales du droit privé*. Paris, 1912.

Durkheim, Émile. *Leçons de sociologie physique des moeurs et du droit*. Paris: Presses Universitaires de France, 1950.  
Translated by Cornelia Brookfield as *Professional Ethics and Civic Morals*. London: Routledge, 1957.

Renner, Karl. *Die Rechtsinstitute des Privatrechts und ihre soziale Funktion*. Tübingen, 1929. Translated by Anges Schwarzschild as *The Institutions of Private Law*, edited by Otto Kahn-Freund. London, 1949.

#### ANALYSIS OF OWNERSHIP

Honoré, A. M. “Ownership.” In *Oxford Essays in Jurisprudence*, edited by A. G. Guest. London: Oxford University Press, 1961.

Noyes, C. Reinold. *The Institution of Property*. New York: Longmans, Green, 1936.

Olivecrona, Karl. *Law as Fact*. London: Oxford University Press, 1939.

Pollock, Frederick. *Jurisprudence and Legal Essays*. London: Macmillan, 1961.

Ross, Alf. *On Law and Justice*. London: Stevens, 1958.

Simpson, A. W. B. “The Analysis of Legal Concepts.” *Law Quarterly Review* 80 (October 1964): 535–558.

*Stanley I. Benn (1967)*

## PROPERTY [ADDENDUM]

What is property? It is some valued item that belongs to someone. Its existence in society may be collective or individual, although even if collective, it usually emerges from instances of (pooled or expropriated) individual

ownership. And that presupposes the right to *private* property.

### PROPERTY IS PRIVATE

The institution of the right to private property is the single most important condition for a society in which freedom in the classical liberal tradition—which means negative liberty, including free trade, freedom of the press, and freedom of religion—is to flourish. Under communism, in contrast, no such right is recognized. Privacy has a negligible role in a system which holds, as Karl Marx (1818–1883) proclaimed, that “the human essence is the true collectivity of man” (1970, p. 126). Even within noncommunist, nonsocialist systems the exact status of property is in dispute—some hold it is a convention established by implicit consensus and maintained by government or law. Some hold it is a natural normative relationship that comes about by means of the creative and productive initiative of persons and the law of property exists to recognize and not to create it.

### WHAT IS THE RIGHT TO PRIVATE PROPERTY?

Karl Marx understood the right to private property, although it was John Locke (1632–1704) who tried to justify this right. Marx wrote, in “On the Jewish Question,” that “the right of man to property is the right to enjoy his possessions and dispose of the same arbitrarily without regard for other men, independently, from society, the right of selfishness” (1970, p. 53). This, though correct, is not the full story. The right to private property, be it applied to obtaining and holding a toothbrush or, as was Marx’s concern (and what Marx found objectionable), an entire factory, does spell out a person’s authority to use what he or she owns without regard for other persons. This use may be reckless, prudent, or generous. Its exercise may not, however, violate others’ rights. Defenders do not assume that it would be insidious.

The natural right to private property was only discussed in direct terms starting in the eleventh and twelfth centuries. William of Ockham (1285–1347) proposed that “Natural right is nothing other than a power to conform to right reason, without an agreement or pact” (2001, p. 48) or, as Heinrich A. Rommen paraphrased him, “the right to private property is a dictate of right reason” (1954, p. 419), the power to make one’s moral choices on one’s own, free of others’ intrusion. Because such choices are made by persons in the natural world, one of our natural rights is the right to private property.

### ONE ROLE OF PRIVATE PROPERTY IN SOCIETY

Property rights weren’t explicitly identified in ancient times but the Old Testament ban on stealing implies what was spelled out by Locke and other classical liberals. Moreover, there have been strong philosophical intimations of it in, for example, Aristotle’s *Politics* (384–322 BCE). Whereas Plato, his teacher, held that, at least within the ruling class of a political community, there may not be any private property and indeed privacy, at all, Aristotle objected as follows:

That all persons call the same thing mine in the sense in which each does so may be a fine thing, but it is impracticable; or if the words are taken in the other sense, such a unity in no way conduces to harmony. And there is another objection to the proposal. For that which is common to the greatest number has the least care bestowed upon it. Every one thinks chiefly of his own, hardly at all of the common interest; and only when he is himself concerned as an individual. For besides other considerations, everybody is more inclined to neglect the duty which he expects another to fulfill; as in families many attendants are often less useful than a few. (*Politics*, 1261b34)

Earlier Thucydides (c. 471–c. 400 BCE) said,

They devote a small fraction of the time to the consideration of any public object, most of it to the prosecution of their own objects. Meanwhile, each fancies that no harm will come to his neglect, that it is the business of somebody else to look after this or that for him; and so, by the same notion being entertained by all separately, the common cause imperceptibly decays. (*The History of the Peloponnesian War*, bk. 1, sec. 141)

So, communal ownership leads to reduction of responsibility and a corresponding lack of attentive involvement with whatever is owned. This does not mean that people are evil. At their homes, this is likely to be different—if one is late and rushes off, the trash will be disposed of upon one’s return. At a public place the attitude seems to be, “It will get cleaned up somehow, by someone, at some time.” So, it is a systemic problem: people are unable to incorporate the significance of managing the public property within the scale of their values. Each of us knows, directly, how important or not it is for oneself to keep one’s backyard clean. So one will take care of it commensurate with that knowledge. It is not possible, however, for an individual to know how important it is for the

community, society, or humanity at large that one keep the air or river or lake clean, and to what degree.

A more recent defense of the right to private property is closer to that which we get from John Locke; namely, that we require this right so as to have a sphere of moral authority—as Robert Nozick (1938–2002) called it, “moral space,” or as Ayn Rand (1905–1982) noted,

Bear in mind that the right to property is a right to an action, like all others: it is not the right to an object, but to the action and the consequences of producing or causing that object. It is not a guarantee that a man will earn any property, but only a guarantee that he will own it if he earns it. It is the right to gain, to keep, to use and to dispose of material values. (1967, p. 322)

Basically, then, the main normative reason given for why one has a right to private property is that it is the means by which one’s liberty to act free of others’ imposition is secured within a social context. It is also a precondition for individuals to act prudently and productively in human communities without the legal permission for others to take from them what they have earned. Economists tend, in contrast, to defend it as a feature of the infrastructure by which productivity and prosperity is best encouraged in a society. Another support given to the idea is that it makes it possible for individuals to remain sovereign and to distribute resources as they see fit rather than others would demand.

There are innumerable objections to the right to private property, most recently the idea that property is held by the public at large and government merely permits individuals to make use of it to the extent government deems this in the public interest. For why this is a troublesome view the general theory of natural rights would need to be explored and scrutinized.

**See also** Civil Disobedience; Cosmopolitanism; Postcolonialism; Republicanism.

### **Bibliography**

- Aristotle. *Politics*. Chicago: University of Chicago Press, 1984.
- Bandman, Elsie and Bertram Bandman, eds. *Bioethics and Human Rights: A Reader for Health Professionals*. Boston: Little, Brown, 1978.
- Comte, August. *The Catechism of Positive Religion*. 1852. Clifton, NJ: Kelley, 1973.
- Locke, John. *The Second Treatise on Government*, edited by Thomas P. Peardon. New York: Liberal Arts Press, 1952.
- Machan, Tibor R. *Individuals and Their Rights*. La Salle, IL: Open Court, 1989.

Machan, Tibor R. *The Right to Private Property*. Stanford, CA: Hoover Institution Press, 2002.

Marx, Karl. *Selected Writings*, edited by David McLellan. London: Oxford University Press, 1970.

Nagel, Thomas, and Liam Murphy. *The Myth of Ownership: Taxes and Justice*. London: Oxford University Press, 2003.

Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books, 1974.

Ockham, William. *Opus Nonaginta Dierum*. Lewiston, NY: E. Mellen Press, 2001.

Rand, Ayn. *Capitalism: The Unknown Ideal*. New York: New American Library, 1967.

Rommen, Heinrich A. “The Genealogy of Natural Rights.” *Thought* 29 (1954).

Thucydides. *The History of the Peloponnesian War*. Mineola, NY: Dover Publications, 2004.

**Tibor R. Machan (2005)**

## PROPOSITIONAL ATTITUDES: ISSUES IN SEMANTICS

Propositional attitudes like knowledge, belief, and assertion play an important foundational role for semantic theory, the goal of which is to specify the meanings of sentences and their semantic contents relative to contexts of utterance. Meanings are plausibly regarded as functions from such contexts to semantic contents, which in turn are closely related to the assertions made, and the beliefs expressed, by utterances. For example, the semantic content of *I live in New Jersey* in a context *C* with *x* as agent and *t* as time is standardly taken to be the proposition that *x* lives in New Jersey at *t*. To understand the meaning of this sentence is, to a first approximation, to know that a competent speaker *x* who sincerely and assertively utters it in *C* asserts, and expresses a belief in, this proposition. Roughly put, if *p* is the semantic content of *S* in *C*, then an assertive utterance of *S* in *C* is an assertion of *p*, and is standardly taken as indicating the speaker’s belief in *p*. Whether the semantic content of a sentence is always among the propositions asserted by an utterance of the sentence, and whether, in those cases in which it is, the assertion of any other proposition by the utterance is always parasitic on the assertion of the semantic content, are matters of detail. Though important, they do not affect the foundational point. A semantic theory for a language is part of a larger theory that interprets the assertions and beliefs of its speakers. This, more than any other fact, allows one to subject semantic theories to empirical test. Competent speakers of a language are relatively good at identifying the propositions

asserted and beliefs expressed by utterances. To the extent to which assignments of semantic content issued by a semantic theory lead to verifiably correct characterizations of speakers' assertions and beliefs, the semantic theory is confirmed; to the extent to which these assignments lead to verifiably incorrect characterizations, it is disconfirmed.

### ATTITUDE ASCRIPTIONS

This point is closely related to the use of attitude ascriptions

(1a) N asserted that S

(1b) N believed that S

to test different semantic analyses of S. It is convenient to express this in terms of the relational nature of the attitudes. Consider assertion. In each case of assertion there is someone, the agent, who does the asserting, and something, the object of assertion, that is asserted. The term *proposition* is used to designate things that are objects of assertion (and other propositional attitudes) and bearers of truth value. Assertion is a mediated relation holding between agents and propositions. An agent asserts a proposition *p* by doing something or employing some content-bearing representation associated with *p*. The most familiar cases are those in which the agent asserts a proposition by assertively uttering a sentence.

Ascriptions like those in (2) report the assertions of agents:

(2a) Edward asserted the proposition that Martha denied.

(2b) Edward asserted the proposition that the Earth is round.

(2c) Edward asserted that the Earth is round.

That in (2a) *asserted* is flanked by two noun phrases suggests that it is a two-place predicate and that a sentence 'NP assert NP' is true if and only if the first (subject) noun phrase designates an agent who bears the assertion relation to the entity designated by the second (direct object) noun phrase. This analysis also applies to (2b), which is true if and only if Edward asserted the proposition designated by *the proposition that the Earth is round*. On the assumption that this proposition is also designated by *that the Earth is round*, this analysis can be extended to (2c), which is equivalent to (2b). Similar remarks hold for other propositional attitude verbs, including *believe*, *deny*, *refute*, and *prove*.

With this in mind, one can return to the ascriptions in (1). If, as many theorists believe, (i) 'that S' in (1) designates the semantic content of S (in the context), (ii) these ascriptions report relations between agents and those contents, and (iii) sometimes substitution of sentences with necessarily equivalent semantic contents fails to preserve the truth values of such ascriptions, then semantic contents must be more fine-grained than the sets of possible world-states in which they are true. On these assumptions substitution in such ascriptions can be used to discriminate different but intensionally equivalent semantic analyses of S.

### PROPOSITIONS, POSSIBLE WORLD-STATES, AND TRUTH SUPPORTING CIRCUMSTANCES

This has significance for possible world semantics. In this framework a semantic theory is a formal specification of truth with respect to a possible context of utterance and circumstance of evaluation. The semantic content of S in C is the set of possible circumstances E such that S is true with respect to C and E. Circumstances of evaluation are traditionally identified with possible world-states—thought of as maximally complete properties that the world genuinely could have had. As a result, the semantic contents of all necessarily equivalent sentences are taken to be identical. This, plus the standard treatment of attitude ascriptions as reporting relations between agents and the semantic contents of their complement clauses, leads to the counterintuitive prediction that substitution of necessarily equivalent sentences in such ascriptions never changes truth value. If one adds the apparently obvious fact that (3a) entails (3b),

(3a) A asserts/believes that P&Q

(3b) A asserts/believes that P&A asserts/believes that Q

one gets the further counterintuitive results (i) that anyone who asserts or believes a proposition *p* asserts or believes all necessary consequences of *p*, and (ii) that no one ever asserts or believes anything necessarily false, since to do so would involve simultaneously asserting or believing every proposition.

In 1983 Jon Barwise and John Perry attempted to evade these results by constructing a semantic theory in which metaphysically possible world-states were replaced by abstract situations—thought of as properties that need be neither maximally complete, nor genuinely capable of being instantiated by any parts of the world. This strategy was shown to be unsuccessful by Scott Soames (1987),

where it was demonstrated that variants of the problems posed by attitude ascriptions for standard possible worlds semantics can be re-created for any choice of truth-supporting circumstances used in formal characterizations of truth with respect to a context and a circumstance. Robert C. Stalnaker explored a different approach in 1984. After providing a naturalistic argument that semantic contents must be sets of metaphysically possible world-states, he suggested that counterexamples could be avoided by (i) allowing for exceptional cases in which attitude ascriptions report relations to propositions other than those expressed by their complement clauses, and (ii) resisting the claim that the agent believes the conjunction of *p* and *q* in many cases in which the agent believes both conjuncts. These suggestions are rebutted in later work by Mark Richard (1990) and Jeffrey Speaks (forthcoming).

### STRUCTURED PROPOSITIONS, MILLIANISM, AND DESCRIPTIVISM

The problems posed by attitude ascriptions for possible worlds semantics have led many theorists to characterize the semantic content of a sentence *S* as a structured complex the constituents of which are the semantic contents of the semantically significant constituents of *S*. In essence this was also the classical position of Gottlob Frege (1892/1948) and Bertrand Russell (1905, 1910). A variant of this position, growing out of the possible worlds framework, was championed by Rudolf Carnap (1947). For Carnap, two formulas are intensionally isomorphic if and only if they are constructed in syntactically the same way from constituents with the same intensions (functions from world-states to extensions). In effect, semantic contents of syntactically simple expressions are identified with intensions, while semantic contents of syntactically complex expressions are structured complexes the constituents of which are the semantic contents of their grammatically significant parts. This view was criticized by Alonzo Church (1954), who argued that semantically complex, but syntactically simple, expressions require a stronger notion of synonymy than sameness of intension. Church's modification of Carnap—which relies on rules of sense to induce a notion of synonymous isomorphism—is a variant of the classical Fregean position.

In the late 1980s the assignment of structured semantic contents to sentences was given a neo-Russellian twist by David Kaplan (1986, 1989), Nathan Salmon (1986), and Soames (1987). On the Russellian picture structured propositions are recursively assigned to formulas, relative

to contexts and assignments of values to variables. The semantic content of a variable *v* relative to an assignment *f* is just *f*(*v*), and the semantic content of a closed (directly referential) term relative to a context *C* is its referent relative to *C*. Semantic contents of *n*-place predicates are *n*-place properties and relations. The contents of truth-functional operators may be taken to be truth functions, while the semantic content of a formula  $\lceil \lambda x [Fx] \rceil$  is identified with a propositional function *g* that assigns to any object *o* the structured proposition expressed by  $\lceil Fx \rceil$  relative to an assignment of *o* to "*x*."  $\lceil \exists x [Fx] \rceil$  expresses the structured proposition in which the property of assigning a true proposition to at least one object is predicated of *g*. In this framework the attitude ascriptions (1a and 1b) express structured semantic contents in which the relation of asserting or believing is predicated of a pair consisting of an agent and the structured proposition semantically expressed by *S*. The semantic theory is completed by specifying the intensions determined by structured semantic contents, including the truth conditions of structured propositions in all possible world-states.

The signature commitment of this approach is to the possibility of asserting and believing singular propositions—which include as constituents the very objects they are about. On this approach to believe *de re* of an object that it is *F* is to believe the singular proposition about that object, which says that it is *F*. Sentences like (4), involving quantifying-in, are quintessential examples of *de re* belief ascriptions.

- (4) There is a planet *x* such that when the ancients saw *x* in the morning they believed that *x* was visible only in the morning and when they saw *x* in the evening they believed that *x* was visible only in the evening.

**MILLIANISM.** If, as Kaplan (1989) contends, the semantic contents of sentences containing indexicals are also singular propositions, then belief ascriptions containing indexicals in their complement clauses are also *de re* and hence share the basic semantic properties of ascriptions like (4). Salmon (1986) and Soames (2002) take this a step further, arguing for the Millian view that the semantic content of an ordinary proper name is simply its referent. One potentially problematic consequence of this view is that since *Ruth Barcan* and *Ruth Marcus* are coreferential, (5a) is characterized as semantically expressing the same proposition as (5b) and hence as having the same semantically determined truth value, even though it seems evident to many that it is possible to believe that



Ruth Barcan was a modal logician without believing the Ruth Marcus was:

(5a) John believes that Ruth Barcan was a modal logician.

(5b) John believes that Ruth Marcus was a modal logician.

Different Millians respond to this problem in different ways. Salmon (1986) and David Braun (2002) argue that the intuitions that (5a) and (5b) can differ in truth value are mistaken because speakers tend to confuse the identical beliefs ascribed to John by these ascriptions with the different manners of holding these beliefs associated with their different sentential complements. Soames (2002, 2005a) argues that even though the semantic contents of these sentences are the same, assertive utterances of them may indeed result in assertions of propositions with different truth values. In “A Puzzle about Belief” (1979) Saul Kripke takes a different tack. While neither advocating nor denying the Millian view, he argues that substitutivity problems of the sort illustrated here are independent of Millianism and indicate a breakdown of the basic principles underlying our belief-reporting practices.

**DESCRIPTIVISM.** By contrast, descriptivists, following in Frege’s footsteps, have wanted to assign different semantic contents to the two names and hence to the complement clauses in (5a and 5b). The problem has been to find a way of doing this that does not run afoul of Kripke’s refutation of descriptivism in *Naming and Necessity* (1972). One of Kripke’s arguments holds that since names are rigid designators, their semantic contents cannot be given by any nonrigid descriptions. This argument is not easily avoided by rigidifying candidate descriptions. As shown by Soames (2002), an analysis that takes the semantic content of *Aristotle* to be given by ‘the actual F’ will, all other things being equal, identify the semantic content of *Aristotle was a philosopher* with the singular proposition (about the actual world-state @) that the unique individual who “was F” in @ was also a philosopher. Assuming that the analysis also includes the standard relational treatment of belief ascriptions, one then gets the result that for any possible agent *a* and world-state *w*, ‘*x* believes that Aristotle was a philosopher’ will be true of *a* with respect to *w* only if in *w* *a* believes that the unique individual who “was F” in @ [not *w*] was also a philosopher. Since this is obviously incorrect, names can neither be nonrigid descriptions, nor descriptions rigidified using the actuality operator.

What about descriptions rigidified using Kaplan’s *dthat* operator? Even if, contra Kripke, a correct reference-fixing description ‘the *x*: *Dx*’ could be found for each name, the semantic content of ‘*dthat* [the *x*: *Dx*]’ would simply be its referent, in which case the descriptivist would be saddled with precisely the Millian predictions about attitude ascriptions that the theory was designed to avoid. One possible response, suggested by David Chalmers (2002), is, in effect, to take a belief ascription ‘*a* believes that *S*’ to report that the belief relation holds between the agent and pair consisting of the semantic content of *S* (in the context) and the meaning (function from contexts to such contents) of *S*. However, now a different problem arises. To avoid Kripke’s non-modal arguments against familiar candidates for reference-fixing descriptions, post-Kripkean descriptivists have had to resort to egocentric, metalinguistic descriptions of the sort *the individual I have heard of under the name “n.”* Although this move assigns different objects of belief to the complement clauses of (5a) and (5b), it does not solve the problem. The point, after all, is not simply to assign different belief objects in these cases, but to explain the different information one gathers about John from utterances of (5a) and (5b). As Soames (2005b) argues in *Reference and Description: The Case against Two-Dimensionalism* (2005), it is hard to see how these egocentric, metalinguistic descriptions could, realistically, contribute to this.

## DAVIDSON’S LINGUISTIC VIEW

A different approach to problems involving substitutivity is to take ascriptions ‘*x* says/asserts/believes that *S*’ as reporting relations either to *S* itself, or to a complex in which *S* is paired with its semantic content. Either way, since substitution of one expression for another in *S* always produces a new complement *S*’, attitude ascriptions that differ in this way always report relations to different objects, whether or not the semantic contents of *S* and *S*’ are the same. This encourages the thought that such ascriptions can always differ in truth value.

An early and influential version of this approach was developed by Donald Davidson (1968–1969), who argued that (6a) should be understood on the model of (6b), in which *that* is treated as a demonstrative, utterances of which refer to utterances of the independent sentence that follows it:

(6a) Galileo said that the Earth moves.

(6b) Galileo said that. The Earth moves.

On this analysis what is said by an assertive utterance *u* of (6a) is that one of Galileo's utterances stands in, as Davidson puts it, the samesaying relation to the subutterance *u\** of *the Earth moves*. Although this analysis promised a simple way of capturing the logic of attitude ascriptions, it foundered on certain recalcitrant facts, including the fact that some ascriptions, like *Every mother said that her son was lovable*, cannot be broken up into separate and independent sentences in the manner of (6b) and the fact that the assertion made by an utterance of (6a) could have been true even if the subutterance *u\** had never existed, indicating that the Davidsonian truth conditions are incorrect.

### LATER LINGUISTIC AND REPRESENTATIONALIST VIEWS

Beginning in the 1990s improvements of Davidson's idea, including, most notably, that of Richard Larson and Peter Ludlow (1993), avoid these difficulties by dispensing with utterances and by treating attitude ascriptions as reporting relations between agents and the interpreted logical forms of their sentential complements. These are abstract, syntax-encoding structures that contain both the expressions occurring in sentences and their referential contents. Abstracting, one has here a version of the structured propositions approach in which linguistic expressions are included in the propositions sentences express. Although this version has potential virtues, it shares a crucial problem with Davidson's original analysis. Just as Davidson's silence about the intension of the samesaying relation prevented his theory from making any predictions about when (if ever) substitution of coreferential names or indexicals in a *says that* ascription changes truth value, so Larson and Ludlow's silence about the intension of the belief relation, alleged to hold between agents and interpreted logical forms, prevents their theory from making any predications about similar substitution in belief ascriptions (see Soames 2002). Since some such substitution clearly does preserve truth value, the problem is a daunting one.

Arguably, the most sophisticated approach of this general type is Richard's (1990), which combines context-sensitivity with linguistically augmented, structured Russellian propositions. For Richard, a belief ascription 'x believes that S', used in a context C, is true of an agent *a* if *a* accepts some sentence *S'* with the same Russellian content in *a*'s context as *S* has in C, while being similar enough to *S* to satisfy the belief-reporting standards in C. As indicated by Soames (2002), the evaluation of this view crucially depends on identifying similarity stan-

dards present in contexts and assessing their impact. Although there are certain evident problems here, opinions of their import vary. Finally, a different sort of context-sensitive view, advocated by Mark Crimmins and John Perry (1989), takes belief ascriptions to report that an agent believes a structured, Russellian proposition by virtue of having ideas of a certain sort—where these are mental particulars in the mind of the agent that are either implicitly demonstrated, or implicitly characterized as being of a certain type, by the one uttering the ascription. This view is usefully criticized by Jennifer Saul (1993).

### EXTENSION: INTENSIONAL "TRANSITIVE" VERBS

Example (2a), in which *assert* occurs as an ordinary transitive verb operating on the extensions of its noun-phrase arguments, shows that not all attitude ascriptions contain sentential clauses. The examples in (7) show that there are also verbs, the grammatical objects of which are not overtly clausal, which are intensional in nature:

- (7a) John wants a perpetual motion machine.
- (7b) John is looking for the fountain of youth.
- (7c) John imagined a room full of unicorns.
- (7d) John worships many gods.

The relationship between these examples and ordinary propositional attitude ascriptions is a matter of ongoing investigation. How is it that (7a to 7d) can be true even though there are apparently no real entities described by their postverbal arguments? Are some or all these sentences covertly clausal? For example, are (7a) and (7b) to be assimilated to (8a) and (8b)?

- (8a) John wants it to be the case that he has a perpetual motion machine.
- (8b) John is trying to bring it about that he finds the fountain of youth.

These and related questions have been discussed by philosophical logicians and linguistic semanticists including Richard Montague (1974), Graeme Forbes (2000), Richard (1998), and Marcel den Dikken, Larson, and Ludlow (1997).

*See also* Intensional Transitive Verbs.

### Bibliography

Barwise, John, and John Perry. *Situations and Attitudes*. Cambridge, MA: MIT Press, 1983.

- Braun, David. "Cognitive Significance, Attitude Ascriptions, and Ways of Believing Propositions." *Philosophical Studies* 108 (2002): 65–81.
- Carnap, Rudolf. *Meaning and Necessity*. Chicago: University of Chicago Press, 1947.
- Chalmers, David. "The Components of Content." In *The Philosophy of Mind: Classical and Contemporary Readings*. New York: Oxford University Press, 2002.
- Church, Alonzo. "Intensional Isomorphism and Identity of Belief." *Philosophical Studies* 5 (1954): 65–73.
- Crimmins, Mark, and John Perry. "The Prince and the Phone Booth: Reporting Puzzling Beliefs." *Journal of Philosophy* 86 (1989): 685–711.
- Davidson, Donald. "On Saying That." *Synthese* 19 (1968–1969): 130–146.
- Dikken, Marcel den, Richard Larson, and Peter Ludlow. "Intensional 'Transitive' Verbs and Concealed Complement Clauses." In *Readings in the Philosophy of Language*, edited by Peter Ludlow, 1041–1053. Cambridge, MA: MIT Press, 1997.
- Forbes, Graeme. "Objectual Attitudes." *Linguistics and Philosophy* 23 (2000): 141–183.
- Frege, Gottlob. "On Sense and Reference." *Zeitschrift für Philosophie und philosophische Kritik* 100 (1892): 25–50; translated by Max Black. *Philosophical Review* 57 (1948): 207–230.
- Kaplan, David. "Demonstratives." In *Themes from Kaplan*, edited by Joseph Almog, John Perry, and Howard Wettstein. New York: Oxford University Press, 1989.
- Kaplan, David. "Opacity." In *The Philosophy of W. V. Quine*, edited by Lewis Edwin Hahn and Paul Arthur Schilpp. La Salle, IL: Open Court, 1986.
- Kripke, Saul. *Naming and Necessity*. Cambridge, MA: Harvard University Press, 1980; originally published in *Semantics of Natural Language*, edited by Donald Davidson and Gilbert Harman. Dordrecht, Netherlands: Reidel, 1972.
- Kripke, Saul. "A Puzzle about Belief." In *Meaning and Use: Papers Presented at the Second Jerusalem Philosophical Encounter, April 1976*, edited by Avishai Margalit. Dordrecht, Netherlands: Reidel, 1979.
- Larson, Richard, and Peter Ludlow. "Interpreted Logical Forms." *Synthese* 95 (1993): 305–356.
- Montague, Richard. *Formal Philosophy*, edited by Richmond H. Thomason. New Haven, CT: Yale University Press, 1974.
- Richard, Mark. "Commitment." *Philosophical Perspectives* 12 (1998): 255–281.
- Richard, Mark. *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them*. Cambridge, U.K.: Cambridge University Press, 1990.
- Russell, Bertrand. "Knowledge by Acquaintance and Knowledge by Description." *Proceedings of the Aristotelian Society* 11 (1910): 108–128.
- Russell, Bertrand. "On Denoting." *Mind* 14 (1905): 479–493.
- Salmon, Nathan. *Frege's Puzzle*. Cambridge, MA: MIT Press, 1986.
- Saul, Jennifer. "Still an Attitude Problem." *Linguistics and Philosophy* 16 (1993): 423–435.
- Soames, Scott. *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*. New York: Oxford University Press, 2002.
- Soames, Scott. "Direct Reference, Propositional Attitudes, and Semantic Content." *Philosophical Topics* 15 (1) (1987): 47–87.
- Soames, Scott. "Naming and Asserting." In *Semantics vs. Pragmatics*, edited by Zoltán Gendler Szabó. New York: Oxford University Press, 2005a.
- Soames, Scott. *Reference and Description: The Case against Two-Dimensionalism*. Princeton, NJ: Princeton University Press, 2005b.
- Speaks, Jeffrey. "The Priority of Belief: Stalnaker on Intentionality." "Is Mental Content Prior to Linguistic Meaning?" *Nous*. Forthcoming.
- Stalnaker, Robert C. *Inquiry*. Cambridge, MA: MIT Press, 1984.

Scott Soames (2005)

## PROPOSITIONAL ATTITUDES: ISSUES IN THE PHILOSOPHY OF MIND AND PSYCHOLOGY

This entry aims to characterize the philosophical issues surrounding the propositional attitudes. Particular attention is paid to the arguments philosophers have brought to bear when discussing the existence and nature of the attitudes.

### SUBJECT MATTER AND PHILOSOPHICAL METHODOLOGY

Discussions of the nature of mind typically distinguish between two fundamental kinds of mental states or properties. One kind of mental state or property involves states that are qualitative in nature: Examples include raw feels, sensations, tickles, and pains. The other kind of mental state or property involves states that are contentful in nature, "pointing to" or "representing" things beyond themselves: Examples include thoughts, desires, fears, and intentions. This distinction is not unproblematic, since it is not clear whether these two categories exhaust the domain of the mental, nor is it clear whether they are mutually exclusive. However, most philosophers of mind accept that there is some important distinction in this region. Propositional attitudes are often cited as the paradigmatic example of the latter kind of mental state.

As their name indicates, the propositional attitudes are *attitudes*—cognitive relations such as belief, desire, fear, hope—that a subject bears to what are typically (though not uncontroversially) taken to be *propositions*. The attitudinal component of a propositional attitude is

a matter of how a particular proposition is being taken: Thus Sally can *believe* that it will rain, *hope* that it will rain, *fear* that it will rain, and so forth. In each case the proposition that is the content of her attitude, *that it will rain*, is the same; what differs is how this proposition is being taken by Sally (believed in, hoped for, or feared). Of course, one and the same attitude can be taken towards different propositions: Thus Sam can believe that the Yankees are a great baseball team, believe that Atlanta is hot and muggy, believe that the office of the U.S. presidency has been demeaned, and so forth. In each case Sam's attitude is the same (belief), what differs is the propositions he believes.

A good deal of the attention philosophers have given to the propositional attitudes is devoted to analyzing the sentences used to ascribe the attitudes. Examples of such *attitude-ascribing sentences* include "Jones believes that it's raining," "Smith worries that State University's soccer team will lose," and "McSorley wants State University's soccer team to lose." Indeed, the very idea that propositional attitudes are cognitive relations that subjects bear to propositions (a variant of which is defended in Jerry Fodor's 1978 article, "Propositional Attitudes") is advanced on the basis of the surface grammar of the sentences used to ascribe the attitudes. Thus "Jones believes that it's raining" is naturally read as saying, of Jones (the reference of "Jones"), that he bears the belief-relation (= the reference of "believes") to the proposition *that it's raining* (= the reference of "that it's raining"). Even sentences such as "McSorley wants State University's soccer team to lose," which do not appear to refer to a proposition at all, can be translated (admittedly with some awkwardness) into equivalent sentences that do, or at least appear to, make such a reference: "McSorley wants it to be the case that State University's soccer team loses." Admittedly, though, such a propositionalist formulation may not be possible in all cases of attitude-ascribing sentences. Consider "Williams fears bats" or "Simon loves ice cream."

Whatever their ultimate nature (more on which below), the propositional attitudes themselves have been thought to be extremely important for the study of human behavior. This is seen when we consider how we go about explaining our own and others' behavior in those cases in which the behavior is taken to be intentional (falling in the domain of human action). In such cases, we explain the behavior as the effect of the subject's propositional attitudes. Thus it seems natural to explain why McSorley walked to the refrigerator in terms of her desire for cold water and her belief that cold water is to be

found there; or to explain why Jackson ran away by citing his belief that a dangerous lion was coming his way and his desire not to get attacked. Explanations of this belief-desire sort are used by ordinary folk as we go about trying to predict and explain the actions of our fellows in everyday circumstances.

One philosophical question that arises in this connection concerns the status of such explanations. Suppose, as many philosophers do, that these explanations are sometimes true. What sort of explanation do they offer? Perhaps everyone can agree that they are *rationalizing* explanations, depicting the action in question as rational in light of the subject's corpus of beliefs and desires. But some philosophers hold that, in addition to rationalizing the behavior in question, they also provide a *causal* explanation of it (see Davidson 1963). If so, then the sort of psychology that appeals to the ordinary "folk" explanations of action—what has been termed *folk psychology*—can take its place beside other sciences that seek to characterize the world's causal nexus.

It is noteworthy that the causal-explanatory perspective provides an alternative approach to the nature of the attitudes, one that differs from the approach involving the analysis of attitude-ascribing sentences. Where the sententialist approach (as we might call it) assumes that we can understand the attitudes by making sense of *our talk* about them, the causal-explanatory approach begins by assuming that, whatever their ultimate nature, the propositional attitudes are the causal springs of human action. Taking the latter approach leads one to conceive of the attitudes as whatever plays the relevant causal role in the production of action. Of course, the two approaches might well be complementary: What one learns about the attitudes from analyzing attitude-ascribing sentences might be compatible with (and supplement) what one learns about the attitudes by thinking about them as the causal basis of action. (Indeed, the desire to secure the compatibility of the sententialist approach and the causal-explanatory approach appears to be a core motivation behind Fodor's 1975 hypothesis in *Language of Thought*, according to which propositional attitudes are tokenings of language-like mental symbols in the brain.) But it is also possible that the sententialist and causal-explanatory approaches will turn out to be in tension, with each one yielding some conclusions not sanctioned by, or perhaps even in conflict with, the other. Settling such a matter is perhaps the main burden of philosophical reflection on the nature of the propositional attitudes.

## THE NATURE OF THE CONTENTS OF THE ATTITUDES

Common to both the sententialist and the causal-explanatory approaches to the propositional attitudes is the idea that the attitudes are *contentful* mental states. As mental states they are *about* things, typically objects and properties from the nonmental environment. Take Sanchez's belief that his grandmother smothers him with kisses. This belief is about his grandmother, him, and the property of smothering with kisses. This aspect of the propositional attitudes—their being *about* worldly states of affairs—raises a number of interesting and related philosophical issues. How does something (such as a mental state like Sanchez's belief) come to be about another thing (such as Sanchez's grandmother) in the first place? What *determines* what a mental state is about? How does the “aboutness” of mental states relate to other forms of “aboutness”? And finally, what can be said about cases in which a mental state is “about” something that does not really exist—unicorns, for example?

Philosophers have introduced the term “intentionality” to designate the domain of aboutness itself. In speaking of mental states as about the world, we are speaking of their *intentional* properties, just as in speaking of, for example, the sentence “Morty Morris has a big red wart on his nose” as about Morty Morris's big red nose wart, we are speaking of the sentence's intentional properties. Such properties are also called *semantic* properties: Both mental states and sentences—and arguably pictures, maps, models, and perhaps other things as well—have such properties. When something, such as a mental state or a sentence, has intentional or semantic properties, and so is about something, we can speak of what the state is about as *the content* of that state. Talk of the content of a mental state is to be understood in terms of what the mental state *represents* as the case. So Sanchez's belief has a content, which is what that belief represents to be the case: namely, that his grandmother smothers him with kisses.

It is noteworthy that a belief can represent something that is not the case. Suppose that Sanchez's grandmother does *not*, in fact, smother him with kisses (it's all “in his head,” so to speak). Then, supposing there is an inventory of all of the facts that make up our world, we would not find in this inventory any fact to the effect that Sanchez's grandmother smothers him with kisses. In short, there is no fact that is represented by his belief. But then what is this shadowy thing we are calling the content of his belief, that which his belief represents to be the case? Above we called this content a “proposition,” and we can now see

the attraction of the view that the content of an attitude is a proposition. For although it is hard to say exactly what propositions are, we can say at least this much: The existence of a given proposition does not depend on the existence of the corresponding fact that would make the proposition true. That is, there can be *false* propositions. Given that Sanchez's grandmother is not as Sanchez's belief depicts her, the proposition that is the content of Sanchez's belief is itself a false proposition.

The postulation of the proposition as the content of the attitudes raises a bundle of related metaphysical questions. What is the nature of propositions? (Is it essentially a linguistic entity? an abstract one? a mental one?) Do propositions have parts, and if so, what is the nature of those parts? Here we focus on a question bearing more directly on the philosophy of mind: How do propositional attitudes come to have the propositional content they have? More concretely, what makes Sanchez's belief a belief about his grandmother, and not, say, about ice cream sundaes or pink elephants or any of an infinite number of other things? Let us address this by asking which facts *fix* the content of his belief: Which facts are such that, if you fix them, then, no matter what else is going on in the universe, you have fixed the content of his belief that his grandmother smothers him with kisses? A natural first guess would be that the facts in question are facts regarding the mental image(s) in Sanchez's mind at the time that he calls this belief to mind. On such a view, once we fix the mental image(s) “in” his mind, we have fixed what his belief is about.

But this cannot be quite right. First, mental images do not appear to have the right sort of specificity to fix the content of the propositional attitudes. To see this, imagine a scenario in which Sanchez's grandmother has an identical twin, from whom the grandmother herself is indistinguishable, but whom Sanchez has never met or otherwise heard of. Then the image in Sanchez's mind “fits” his grandmother's twin as much as it “fits” his grandmother. But it seems implausible to think that his belief is about the twin, for he has never met or heard of her. Second, in addition to not having the right sort of specificity, mental images are too unstable and subjective to fix the contents of one's attitudes. This is clearest in cases in which the subject matter of the attitude is an abstract one. Precisely what image goes before your mind when you call forth your belief that  $1+1=2$ ? And what image is before your mind when you believe that space is (or is not) infinite? Will it be true that any two people who believe e.g. that  $1+1=2$  will have the same type of image before their minds? Presumably not. But then how

does the image fix the content of their belief? It seems that what they have in common, in virtue of which they both count as believing that  $1+1=2$ , is something other than a particular type of image.

And the same point can be made even in cases in which the subject matter of the attitude is not abstract. Take Roger's belief that Morty Morris has a big red wart on his nose. Since Morty Morris is Roger's best friend, Roger has a vivid mental image of Morty (wart and all). But Mathilde, who (having been told by Roger) also believes that Morty Morris has a big red wart on his nose, has never met Morty, and therefore has no such image. Again it seems that what Roger and Mathilde have in common, in virtue of which they both count as believing that Morty Morris has a big red wart on his nose, is something other than a particular type of mental image in mind.

These arguments (and the examples on which they are based) raise a host of issues regarding how propositional attitudes come to have the propositional content they have. Consider first the relation between such contents and the environment in which one lives and interacts. One plausible account of why Sanchez's belief is about his grandmother, rather than her identical twin, is that his belief was caused and sustained by activities involving one woman and not the other. So it can seem that interaction with one's environment is relevant to the determination of the contents of one's attitudes. Next consider the relationship between language and the content of the attitudes. Recalling that mental images are too unstable and subjective to fix the contents of attitudes, we might ask: Precisely what do Roger and Mathilde have in common, in virtue of which they both count as believing that Morty Morris has a big red wart on his nose? At least part (but only part!) of the answer is that they are both disposed to accept and assert a sentence that *means* that Morty Morris has a big red wart on his nose. Perhaps, then, among the facts that fix the content of one's attitudes we must include facts regarding the meanings of one's words.

These conclusions highlight one of the bigger controversies in the theory of content. In particular, we have seen at least three types of fact that might be regarded as relevant to fixing the content of one's attitudes. We started off with the suggestion that facts regarding the subject's mental images fix the content of her attitudes, but we moved quickly to include facts regarding causal history and then on to facts regarding the meanings of one's words. These correspond roughly to three distinct

theoretical options available with respect to the sort of facts needed to fix the content of one's attitudes.

*Content internalism* is the view that the only facts needed to fix the content of a subject's attitudes are facts that do not presuppose the existence of anything beyond the subject herself. The view with which we started, according to which the facts regarding the subject's mental images fix the contents of her attitudes, is one version of content internalism. But the content internalist can allow other sorts of facts, so long as these do not presuppose the existence of anything beyond the subject herself; and the most plausible versions of content internalism (for which see Searle 1983) include facts about the individual's use of language, where the meanings of her words are not thought to depend on the existence of anything beyond the subject herself. Of the various arguments for content internalism, one of the most influential is what we might call the argument from "intentional inexistence." Consider, to begin, that one can form a belief which is "about" something that does not exist—as with Ponce de Leon's belief that the Fountain of Youth is in Florida, or Roger's belief that the largest natural number is even. What is more, it would seem possible (though of course highly unlikely) that none of our beliefs succeed in being about any existing thing: Perhaps you are suffering an eternal and systematic hallucination in a world containing nothing but your own mind! But in that case, although your beliefs remain the same (or so it might seem), there are no worldly objects for them to be "about." This suggests that the "aboutness" properties of beliefs should be understood in such a way as not to presuppose the existence of anything beyond the thinking subject.

Many philosophers, unconvinced by this sort of argument, have thought that the internalist view is too restrictive in the set of facts it regards as relevant to fixing the content of the attitudes. A second view, *content individualism*, expands the set of content-fixing facts to include not just the facts allowed by the content internalist, but also any facts regarding the thinker's own causal history. (See Davidson 1984 and 2001 for an example of a view that combines *content externalism*, which is the denial of content internalism, with content individualism.) Although the cost of moving from internalism to individualism is that of having to rebut the argument from intentional inexistence—something that forces the individualist to come up with an account of beliefs "about" non-existent "objects"—the payoff of making this move can be made clear in connection with the following development of Sanchez's case. Sanchez

has an identical twin, Twin-Sanchez, separated from Sanchez from birth. Twin-Sanchez has interacted only with twin-granny, the identical twin of Sanchez's grandmother. Further, the course of experience Sanchez has with his grandmother is internally indistinguishable from the course of experience Twin-Sanchez has with twin-granny. So, for example, at the very moment Sanchez sees his grandmother wearing a lovely purple vest and making waving motions as she smiles, Twin-Sanchez sees twin-granny wearing an indistinguishable lovely purple vest making waving motions as *she* smiles; at the very moment Sanchez hears his grandmother singing a lovely melody, Twin-Sanchez hears twin-granny singing an indistinguishable lovely melody; and so forth through time. At one point each of the Sanchez twins, admiring the grandmother in his presence, forms a belief he would express with, "She has a wonderful voice." The natural view is that the contents of their beliefs differ: Sanchez's belief represents *his* grandmother (not twin-granny) as having a wonderful voice, whereas twin-Sanchez's belief represents twin-granny (not granny) as having a wonderful voice.

The content individualist can easily accommodate this natural view, as the difference in content can be fixed by the facts regarding each twins' causal history (with distinct grannies). The content internalist, by contrast, will have trouble accepting the natural view: Since the twins' course of experiences are internalistically indistinguishable, there will be some pressure on the content internalist to treat the twins as having beliefs *with the very same content*. (See Searle 1983 for an attempt by an internalist to avoid this conclusion.)

But if the content individualist has this virtue over the internalist, some philosophers have felt that individualism does not go far enough. A third position, which we might designate as *content anti-individualism*, is still more liberal in the range of facts it regards as relevant to fixing the content of a subject's attitudes. As its name suggests, content anti-individualism is the denial of content individualism. But it is helpful to see why a theorist might deny that "individualistic" facts suffice to fix the content of a subject's attitudes. The controversy has to do with the role of language in fixing the content of the attitudes. In one sense, it is uncontroversial that the meaning of one's words determines the contents of the attitudes one expresses with those words. The controversial matter regards what determines the meaning of one's words. The individualist maintains that no facts beyond those regarding the individual speaker herself—the conditions under which she uses her words, how she herself expli-

cates their meanings—are needed to fix the meaning of her words; whereas the anti-individualist maintains that these "individualistic" facts do not suffice to fix the meanings of her words. The insight (or alleged insight) behind anti-individualism is that individual language users typically defer to, and take themselves to be answerable to, public standards of correct usage. Such standards are not typically fixed by the individual's *own* word usage or meaning-explications, but instead are fixed by the usage of other speakers (Kripke 1972) and the meaning-explicating practices of the relevant experts in her linguistic community (Putnam 1975 and Burge 1979).

Interestingly, the sententialist and causal-explanatory approaches to the attitudes bear on the debate regarding the nature of mental content. For example, among the reasons offered in defense of anti-individualism, Burge notes in "Individualism and the Mental" (1979) that variations in public standards for the correct use of a word lead to differences in *the belief-attributing sentences* that would be used to report a subject's beliefs. And among the reasons offered in defense of content internalism are considerations pertaining to the internal basis of mental causes (for which see Fodor 1980). Although neither argument is decisive, each suggests the core motivations for and potential liabilities of the various positions on mental content.

### THE METAPHYSICS OF THE ATTITUDES: VERSIONS OF MATERIALISM

The question regarding the nature of mental content cannot be addressed in isolation from what we might call the metaphysics of the attitudes. What is the nature of the states and properties dubbed "the propositional attitudes"? How do such states and properties relate to the thinker's bodily states and properties? These questions, of course, force us to confront a particular version of the notorious mind-body problem.

The positions that can be taken on the relation between a subject's propositional attitudes and her bodily states and properties correspond to positions familiar from the general mind-body problem. *Attitude dualism* holds that propositional attitudes are *immaterial* states or properties of thinking subjects. But as with dualism generally, attitude dualism runs into trouble in connection with the causal role that the attitudes are thought to play: How do immaterial states or properties affect a subject's body? Most contemporary philosophers take some version of this problem to be decisive against dualism. And of these most go on to endorse *materialism*, according to

which all of the objects and properties of our world are material in nature. So we will restrict our discussion accordingly.

Among materialist views we can begin with the view known as *philosophical behaviorism*, according to which the so-called propositional attitudes really are nothing other than complex behavioral dispositions. Philosophical behaviorism itself (unlike psychological behaviorism) was originally motivated by the verification theory of meaning, according to which the meaning of a sentence consists in the conditions whose obtaining would *verify* the sentence (establish its truth). Since sentences such as “John believes that it’s raining” are typically regarded as true or false in virtue of observable behavior (e.g., John’s uttering “It’s raining!,” carrying an umbrella with him, putting on galoshes, and so forth), the result of applying the verification theory of meaning to attitude-ascribing sentences is that each such sentence is to be regarded as equivalent in meaning to a “behavioral translation,” a much longer sentence describing all of the behaviors and behavioral dispositions whose presence would verify the original sentence (see e.g. Ryle 1949). However, this view faces two obvious and devastating difficulties.

First, as noted in Putnam “Brains and Behavior” (1965), the view is either false or unacceptably circular. It is false if the translation of the target sentence (“John believes that it’s raining”) fails to capture all of the conditions whose presence would be taken as evidence for the truth of that sentence. But in order to avoid falsity on this score, the translation will need to make reference to other attitudes the subject has: For example, John’s uttering “It’s raining!” counts for the truth of “John believes that it is raining” only if he is speaking sincerely and believes that “It’s raining” means that it’s raining; John’s taking an umbrella with him (or putting galoshes on) counts for the truth of “John believes that it’s raining” only if he desires not to get wet and believes that the umbrella (galoshes) will prevent him from getting wet; and so forth. In fact, it would appear that the connection between attitudes and behavior invariably involves other attitudes in this way. But in that case, any attempt to translate a target attitude-ascribing sentence will yield a translation which itself contains mention of other attitudes. On pain of circularity, these latter attitude-ascribing components in the translation must also be translated. But then the problem begins again, and the whole approach appears doomed to an unacceptable sort of circularity.

Nor is this philosophical behaviorism’s only problem. A second objection is that philosophical behaviorism

surrenders the idea of propositional attitudes as *the causes* of behavior. Consider: that a sugar cube dissolves in water is the basis for regarding it as water-soluble; so it would be an empty explanation to regard its solubility in water as *the cause* of its dissolving on a particular occasion. (Compare the doctor spoofed in Molière’s play *La malade imaginaire*: He explained the sleep-producing character of a particular drug to its having a “dormative virtue.”) Similarly, if beliefs and desires are dispositions to act, then it would be an empty explanation indeed to regard beliefs and desires as *the causes* of action.

Given the failure of philosophical behaviorism, the desire to preserve the causal profile of the propositional attitudes within a materialist framework provided the main motivation behind *identity theory*. Recognizing the role of the attitudes as the causes of intelligent behavior, early identity theorists used the fact (or what they regarded as the fact) that the causes of intelligent behavior are to be found in the states and processes of the central nervous system, to conclude that the propositional attitudes are *identical* to those states and processes of the central nervous system. (For an early formulation of identity theory, albeit in connection with sensory rather than contentful states, see Smart 1962.) The proposed identity was between property-*types*: An “attitudinal” property-type (such as the property of *believing that it is raining*) was held to be identical to a property-type instantiated by the central nervous system (such as the property of *having such-and-such a pattern of neural activation in this-or-that region of the brain*). But this gave rise to an objection from the so-called multiple-realizability of mental states (Putnam 1967): On the assumption that creatures whose underlying neurophysiology is very different from our own might nevertheless be regarded as being the subjects of attitudes, such type-identity claims were much too strong.

Such an objection to type-identity theory acquired additional force in light of the development of sophisticated forms of artificial intelligence. Alan Turing’s famous “Turing Test” (1950) taught that a system was to be regarded as “intelligent” so long as it behaved in a way that would lead those with whom it interacted to regard it as intelligent. The implicit idea was that any system with the right sort of *functional complexity*—as seen in its capacity to acquire and process information from its environment and to use this information to guide its subsequent actions—was to be regarded as intelligent (and hence, given some plausible assumptions, as a subject of the propositional attitudes). The result was what is perhaps the most widely accepted view of the metaphysics of



the attitudes: *functionalism*. According to the functionalist, propositional attitudes are best characterized by their functional or causal profile. So just as it would be a mistake to identify the property of *being a carburetor* with the property of *being made of metal and shaped in such-and-such a way*—surely being a carburetor is more a matter of function rather than material—so too it would be a mistake to identify the property of *believing that it is raining* with some particular property of the body. Rather, a subject has this property, and so counts as believing that it is raining, when the subject is in a state with a certain functional or causal profile—one that is caused in certain characteristic sorts of ways (e.g., seeing rain) and interacts with other functionally defined states to bring about certain effects (e.g., producing utterances of “It’s raining!,” movements to retrieve the umbrella when leaving, and so forth).

According to the functionalist, the first task in connection with the metaphysics of the attitudes is to specify the functional role corresponding to each distinct attitude. Once that task is completed, the functionalist philosopher can then pass on to empirical investigation the task of identifying what particular physical property *realizes* that functional role in a given system. (Compare: Once the functional role of a carburetor has been specified, we can go on and ask which feature of a particular car realizes that role.) Such a view is often advanced as part of a “computational” theory of the attitudes, according to which the functional role of particular mental states is best understood in information-processing terms. So formulated, functionalism, as shown by A. Newell (1980) and David Marr (1982), has been popular not only in the philosophy of mind but also in traditional cognitive science.

Of course, having specified what we take to be the functional role of a particular attitude-type (say, the belief that it is raining), there is no guarantee that there will *be* any state or property of the body or the central nervous system playing that role. Perhaps the very idea that there is a state playing that role is itself part of a mistaken theory of the mind, one whose fundamental postulates (beliefs, desires, and so forth) are as misguided as was the postulation of witches and other spiritual entities by misguided theorists of earlier ages. A number of philosophers, such as P. Churchland (1981), have begun to express such misgivings, arguing that the account of mind which postulates propositional attitudes is part of a “worm-eaten myth” that will be replaced as brain science progresses. Such a view, known as *eliminative materialism*, is perhaps the starkest version of materialism there

is, as it combines a general commitment to materialism with the view that there is nothing in the material world that answers to what we take the propositional attitudes to be. Though clearly radical, such a view has challenged mainstream theorists to further clarify what is at issue in the debate over the propositional attitudes.

**See also** Belief; Belief Attributions; Content, Mental; Intentionality; Language of Thought.

### Bibliography

- Burge, Tyler. “Individualism and the Mental.” In *Studies in Metaphysics*, edited by Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein. 73–121. Midwest Studies, vol. 4. Minneapolis: University of Minnesota Press, 1979. Reprinted *The Nature of Mind*, edited by David M. Rosenthal. Oxford: Oxford University Press, 1991.
- Churchland, P. “Eliminative Materialism and the Propositional Attitudes.” *The Journal of Philosophy* 78 (2) (1981): 67–90.
- Davidson, Donald. “Actions, Reasons, and Causes.” [1963] Reprinted in *Essays on Actions and Events*. Oxford: Oxford University Press, 1980.
- Davidson, Donald. *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press, 1984.
- Davidson, Donald. *Subjective, Intersubjective, Objective*. Oxford: Oxford University Press, 2001.
- Fodor, Jerry. *The Language of Thought*. New York: Thomas Y. Crowell, 1975.
- Fodor, Jerry. “Methodological Solipsism considered as a Research Strategy in Cognitive Psychology.” *The Behavioral and Brain Sciences* 3 (1) (1980): 63–72. Reprinted in *The Nature of Mind*, edited by David M. Rosenthal. Oxford: Oxford University Press, 1991.
- Fodor, Jerry. “Propositional Attitudes.” *The Monist* 61 (4) (1978): 501–23. Reprinted in *The Nature of Mind*, edited by David M. Rosenthal. Oxford: Oxford University Press, 1991.
- Kripke, Saul A. *Naming and Necessity*. Cambridge, MA: Harvard University Press, 1980.
- Marr, David. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W. H. Freeman, 1982.
- Newell, A. “Physical Symbol Systems.” *Cognitive Science* 4 (2) (1980): 135–183.
- Putnam, Hilary. “Brains and Behavior.” In *Analytical Philosophy, Second Series*, edited by R. J. Butler. Oxford: Basil Blackwell and Mott, 1965.
- Putnam, Hilary. “The Meaning of ‘Meaning.’” In his *Philosophical Papers*. Vol. 2: *Mind, Language, and Reality*. Cambridge, U.K.: Cambridge University Press, 1975.
- Putnam, Hilary. “The Nature of Mental States.” [1967] Reprinted *The Nature of Mind*, edited by David M. Rosenthal. Oxford: Oxford University Press, 1991.
- Ryle, Gilbert. *The Concept of Mind*. Chicago: University of Chicago Press, 1949.
- Searle, John R. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge, U.K.: Cambridge University Press, 1983.

Smart, J. J. C. "Sensations and Brain Processes." [1962]  
 Reprinted in *The Nature of Mind*, edited by David M.  
 Rosenthal. Oxford: Oxford University Press, 1991.  
 Turing, Alan. "Computing Machinery and Intelligence." *Mind*  
 59 (1950): 433–460.

*Sanford Goldberg (2005)*

## PROPOSITIONAL KNOWLEDGE, DEFINITION OF

The traditional "definition of propositional knowledge," emerging from Plato's *Meno* and *Theaetetus*, proposes that such knowledge—knowledge that something is the case—has three essential components. These components are identified by the view that knowledge is justified true belief. Knowledge, according to the traditional definition, is belief of a special kind, belief that satisfies two necessary conditions: (1) the truth of what is believed and (2) the justification of what is believed. While offering various accounts of the belief condition, the truth condition, and the justification condition for knowledge, many philosophers have held that those three conditions are individually necessary and jointly sufficient for propositional knowledge.

The belief condition requires that one accept, in some manner, any proposition one genuinely knows. This condition thus relates one psychologically to what one knows. It precludes that one knows a proposition while failing to accept that proposition. Some contemporary philosophers reject the belief condition for knowledge, contending that it requires a kind of mentalistic representation absent from many cases of genuine knowledge. Some other contemporary philosophers endorse the belief condition but deny that it requires actual assent to a proposition. They propose that, given the belief condition, a knower need only be disposed to assent to a proposition. Still other philosophers hold that the kind of belief essential to propositional knowledge requires assent to a known proposition, even if the assent need not be current or ongoing. The traditional belief condition is neutral on the exact conditions for belief and for the objects of belief.

The truth condition requires that genuine propositional knowledge be factual, that it represent what is actually the case. This condition precludes, for example, that astronomers before Nicolas Copernicus knew that Earth is flat. Those astronomers may have believed—even justifiably believed—that Earth is flat, as neither belief nor

justifiable belief requires truth. Given the truth condition, however, propositional knowledge without truth is impossible. Some contemporary philosophers reject the truth condition for knowledge, but they are a small minority. Proponents of the truth condition fail to agree on the exact conditions for the kind of truth essential to knowledge. Competing approaches to truth include correspondence, coherence, semantic, and redundancy theories, where the latter theories individually admit of variations. The truth condition for knowledge, generally formulated, does not aim to offer an exact account of truth.

The justification condition for propositional knowledge guarantees that such knowledge is not simply true belief. A true belief may stem just from lucky guesswork; in that case it will not qualify as knowledge. Propositional knowledge requires that the satisfaction of its belief condition be suitably related to the satisfaction of its truth condition. In other words, a knower must have adequate indication that a belief qualifying as knowledge is actually true. This adequate indication, on a traditional view of justification suggested by Plato and Immanuel Kant, is suitable evidence indicating that a proposition is true. True beliefs qualifying as knowledge, on this traditional view, must be based on justifying evidence.

Contemporary philosophers acknowledge that justified contingent beliefs can be false; this is fallibilism about epistemic justification, the kind of justification appropriate to propositional knowledge. Given fallibilism, the truth condition for knowledge is not supplied by the justification condition; justification does not entail truth. Similarly, truth does not entail justification; one can lack evidence for a proposition that is true.

Proponents of the justification condition for knowledge do not share an account of the exact conditions for epistemic justification. Competing accounts include epistemic coherentism, which implies that the justification of any belief depends on that belief's coherence relations to other beliefs, and epistemic foundationalism, which implies that some beliefs are justified independently of any other beliefs. Recently, some philosophers have proposed that knowledge requires not evidence but reliable (or truth-conducive) belief formation and belief sustenance. This is reliabilism about the justification condition for knowledge. Whatever the exact conditions for epistemic justification are, proponents of the justification condition maintain that knowledge is not merely true belief.

Although philosophers have not agreed widely on what specifically the defining components of proposi-

tional knowledge are, there has been considerable agreement that knowledge requires, in general, justified true belief. Traditionally, many philosophers have assumed that justified true belief is sufficient as well as necessary for knowledge. This is a minority position now, owing mainly to Gettier counterexamples to this view. In 1963 Edmund Gettier challenged the view that if one has a justified true belief that  $p$ , then one knows that  $p$ . Gettier's counterexamples are:

- (I) Smith and Jones have applied for the same job. Smith is justified in believing that (i) Jones will get the job, and that (ii) Jones has ten coins in his pocket. On the basis of (i) and (ii), Smith infers, and thus is justified in believing, that (iii) the person who will get the job has ten coins in his pocket. As it turns out, Smith himself will actually get the job, and he also happens to have ten coins in his pocket. So, although Smith is justified in believing the true proposition (iii), Smith does not know (iii).
- (II) Smith is justified in believing the false proposition that (i) Jones owns a Ford. On the basis of (i), Smith infers, and thus is justified in believing, that (ii) either Jones owns a Ford or Brown is in Barcelona. As it turns out, Brown is in Barcelona, and so (ii) is true. So although Smith is justified in believing the true proposition (ii), Smith does not know (ii).

Gettier counterexamples are cases where one has a justified true belief that  $p$  but lacks knowledge that  $p$ . The Gettier problem is the difficulty of finding a modification of, or an alternative to, the traditional justified-true-belief analysis that avoids difficulties from Gettier counterexamples.

Contemporary philosophers have not reached a widely accepted solution to the Gettier problem. Many philosophers take the main lesson of Gettier counterexamples to be that propositional knowledge requires a fourth condition, beyond the justification, belief, and truth conditions. Some philosophers have claimed, in opposition, that Gettier counterexamples are defective because they rely on the false principle that false evidence can justify one's beliefs. There are, however, examples similar to Gettier's that do not rely on any such principle. Here is one such example inspired by Keith Lehrer and Richard Feldman:

- (III) Suppose that Smith knows the following proposition,  $m$ : Jones, whom Smith has always found to be reliable and whom Smith has no reason to dis-

trust now, has told Smith, his officemate, that  $p$ : He, Jones, owns a Ford. Suppose also that Jones has told Smith that  $p$  only because of a state of hypnosis Jones is in and that  $p$  is true only because, unknown to himself, Jones has won a Ford in a lottery since entering the state of hypnosis. Suppose further that Smith deduces from  $m$  its existential generalization,  $o$ : There is someone, whom Smith has always found to be reliable and whom Smith has no reason to distrust now, who has told Smith, his officemate, that he owns a Ford. Smith, then, knows that  $o$ , since he has correctly deduced  $o$  from  $m$ , which he also knows. Suppose, however, that on the basis of his knowledge that  $o$ , Smith believes that  $r$ : Someone in the office owns a Ford. Under these conditions, Smith has justified true belief that  $r$ , knows his evidence for  $r$ , but does not know that  $r$ .

Gettier counterexamples of this sort are especially difficult for attempts to analyze the concept of propositional knowledge.

One noteworthy fourth condition consists of a "defeasibility condition" requiring that the justification appropriate to knowledge be "undefeated" in that an appropriate subjunctive conditional concerning defeaters of justification be true of that justification. A simple defeasibility condition requires of our knowing that  $p$  that there be no true proposition,  $o$ , such that if  $q$  became justified for us,  $p$  would no longer be justified for us. If Smith genuinely knows that Laura removed books from the office, then Smith's coming to believe with justification that Laura's identical twin removed books from the office would not defeat the justification for Smith's belief regarding Laura herself. A different approach claims that propositional knowledge requires justified true belief sustained by the collective totality of actual truths. This approach requires a precise, rather complex account of when justification is defeated and restored.

The importance of the Gettier problem arises from the importance of a precise understanding of the nature, or the essential components, of propositional knowledge. A precise understanding of the nature of propositional knowledge, according to many philosophers, requires a Gettier-resistant account of knowledge.

*See also* Coherentism; Epistemology; Kant, Immanuel; Plato; Reliabilism; Truth.

### **Bibliography**

Audi, Robert. *Epistemology*. London: Routledge, 1998.

- BonJour, L. *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press, 1985.
- Carruthers, P. *Human Knowledge and Human Nature*. New York: Oxford University Press, 1992.
- Chisholm, Roderick. "Knowledge as Justified True Belief." In *The Foundations of Knowing*. Minneapolis: University of Minnesota Press, 1982.
- Chisholm, Roderick. *Theory of Knowledge*. 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 1989.
- Goldman, A. *Epistemology and Cognition*. Cambridge, MA: Harvard University Press, 1986.
- Greco, John, and Ernest Sosa, eds. *The Blackwell Guide to Epistemology*. Oxford: Blackwell, 1998.
- Klein, Peter. "A Proposed Definition of Propositional Knowledge." *Journal of Philosophy* 67 (1971): 471–482.
- Lehrer, K. *Theory of Knowledge*. Boulder, CO: Westview Press, 1990.
- Lewis, C. I. *An Analysis of Knowledge and Valuation*. La Salle, IL: Open Court, 1946.
- Moser, Paul. "Epistemology (1900–Present)." In *Routledge History of Philosophy*. Vol. 10: *Philosophy of the English Speaking World in the 20th Century*, edited by John Canfield. London: Routledge, 1996.
- Moser, Paul. *Knowledge and Evidence*. New York: Cambridge University Press, 1989.
- Moser, Paul, ed. *The Oxford Handbook of Epistemology*. Oxford: Oxford University Press, 2002.
- Moser, Paul, and A. vander Nat, eds. *Human Knowledge: Classical and Contemporary Approaches*. 2nd ed. New York: Oxford University Press, 1995.
- Pollock, J. *Contemporary Theories of Knowledge*. Totowa, NJ: Rowman & Littlefield, 1986.
- Shope, Robert. *The Analysis of Knowing: A Decade of Research*. Princeton, NJ: Princeton University Press, 1983.
- Sosa, E. *Knowledge in Perspective*. New York: Cambridge University Press, 1991.

**Paul K. Moser (1996)**

*Bibliography updated by Benjamin Fiedor (2005)*

## PROPOSITIONS

On one use of the term, "propositions" are objects of assertion, what successful uses of declarative sentences say. As such, they determine truth-values and truth conditions. On a second, they are the objects of certain psychological states (such as belief and wonder) ascribed with verbs that take sentential complements (such as *believe* and *wonder*). On a third use, they are what are (or could be) named by the complements of such verbs. Many assume that propositions in one sense are propositions in the others.

After some decades of skepticism about the worth of positing propositions, the last quarter of the twentieth century saw renewed interest in and vigorous debate over their nature. This can be traced in good part to three fac-

tors: the development in intensional logic of formal models of propositions; (not altogether unrelated) attacks on broadly Fregean accounts of propositions; and a spate of work on the nature of belief and its ascription.

"Possible-worlds semantics" is a collection of methods for describing the semantical and logical properties of expressions such as *necessarily*; these methods developed out of work done by Saul Kripke, Richard Montague, and others in the 1960s. It illuminated the logic and semantics of modal terms such as *necessarily*, of conditionals and tenses, and other constructions as well. In such semantics one assigns a sentence a rule that determines a truth-value relative to various "circumstances of evaluation" (possible worlds, times, whatever); a sentence such as "it is necessary that S" has its truth-value determined by the rule so associated with S. The success of such accounts made it natural to hypothesize that propositions, qua what is named by expressions of the form "that S," could be identified with such rules—equivalently, with sets of circumstances such rules pick out.

Such a conception of proposition provides too crude an account of objects of belief or assertion: It implausibly makes all logically equivalent sentences express the same belief and say the same thing. A partial solution to this problem supposes that propositional identity is partially reflected in sentential structure, taking propositions themselves to be structured. Given the working hypothesis that a proposition's structure is that of sentences expressing it, critical to determining the proposition a sentence (use) expresses are the contributions made by sentence parts (on that use).

Gottlob Frege (1952) suggested that associated with names and other meaningful expressions are "ways of thinking" or *senses* of what the expressions pick out; one might suppose that sense and sentence structure jointly determine proposition expressed. Sense, in the case of names and other singular terms, has standardly been taken to be given by describing how one thinks of the referent. For example, the sense of "Aristotle" for me might be given by "the author of the *Metaphysics*"; if so, my uses of "Aristotle taught Alexander" and "the author of the *Metaphysics* taught Alexander" would, on a Fregean view, express the same proposition.

During the 1970s Kripke, David Kaplan, and others argued convincingly that this view is untenable: It is obvious, on reflection, that the truth conditions of the assertion or belief that Aristotle was *F* depend on Aristotle in a way in which the truth conditions of the assertion or belief that the author of the *Metaphysics* was *F* do not. So either ways of thinking are somehow tied to the objects

they present (so that the way I think of Aristotle could not present anything but Aristotle), or the contributions of expressions to propositions must be something other than senses.

The success of accounts of intensional language that ignored sense in favor of constructions from references, along with the apparent failure of Fregean accounts, led in the 1980s to debate over the merits of what is variously called direct-reference theory, Millianism, and (neo-)Russellianism, espoused at various times by a wide variety of theorists including Kaplan (1989), Mark Richard (1990), Nathan Salmon, and Scott Soames (1988). On such views sense is irrelevant to individuating a proposition; indeed, it is irrelevant to semantics. In particular, what a name contributes to a proposition is its referent: The proposition that Twain is dead is the same singular proposition as the proposition that Clemens is.

Neo-Russellians identify the object of assertion and the referent of a “that” clause with a Russellian proposition. They allow that there is such a thing as a “way of grasping” a proposition and that belief in a singular proposition is mediated by such. Against the intuition that, for example, A: Mo believes that Twain is dead, and B: Mo believes that Clemens is dead, might differ in truth-value, direct-reference accounts typically suggest that a pragmatic explanation is appropriate. Just as an ironic use of a sentence can convey a claim without literally expressing it, so a sentence about Mo’s beliefs might convey information about Mo’s way of grasping a singular proposition, without that information being part of what the sentence literally says. If this is so, intuitions about A and B are explained pragmatically.

Those unhappy with this account of propositions have looked elsewhere. Many accounts of propositions identify the proposition determined by S with some construction from linguistic items associated with S and the semantic values of S’s parts. James Higginbotham has identified the referents of “that” clauses with phrase markers that may be annotated with referents; Richard has suggested that the referent of a “that” clause be identified with something like the singular proposition it determines paired off with the sentence itself. In making linguistic items constitutive of propositions, these views run counter to ones, like Frege’s and Bertrand Russell’s, that closely tie meaning and synonymy to propositional determination. On linguistic views of propositions the synonymy of *groundhog* and *woodchuck* does not assure the identity of the proposition that groundhogs are pests with the proposition that woodchucks are. Other theo-

rists (Gareth Evans, for example) have attempted to revive a version of Frege’s views of propositions.

Many philosophers continue to doubt the utility of positing propositions. Quineans argue that meaning and reference must be determined by behaviorally manifest facts but that such facts woefully underdetermine assignments of meaning and reference; they conclude that there is nothing about language that need or could be explained by positing propositions. Stephen Schiffer has argued that propositions are a sort of “linguistic posit”: that we accept nominalizations of the form “that S” as referring to singular terms and have coherent criteria for using sentences in which those terms occur is itself sufficient for its being true that there are propositions. Such a deflationist view implies neither the possibility of a substantive account of propositions (on which, for example, the proposition expressed by a sentence is compositionally determined), nor that propositions play a substantive role in explaining semantic phenomena.

*See also* Frege, Gottlob; Kripke, Saul; Meaning; Modality, Philosophy and Metaphysics of; Philosophy of Language; Quine, Willard Van Orman; Reference; Russell, Bertrand Arthur William.

### **Bibliography**

- Evans, G. *The Varieties of Reference*. Oxford: Clarendon Press, 1982.
- Frege, G. “Über Sinn und Bedeutung.” *Zeitschrift für Philosophie und Philosophische Kritik* 100 (1892): 25–50. Translated by P. Geach and M. Black as “On Sense and Reference.” In *Translations from the Philosophical Writings of Gottlob Frege*, edited by P. Geach and M. Black. New York: Philosophical Library, 1952.
- Horwich, P. *Truth*. Cambridge, MA: Blackwell, 1990.
- Kaplan, D. “Demonstratives.” In *Themes from Kaplan*, edited by J. Perry et al. New York: Oxford University Press, 1989.
- Kripke, S. *Naming and Necessity*. Cambridge, MA: Harvard University Press, 1980.
- Lewis, D. “Attitudes *de dicto* and *de se*.” *Philosophical Review* 88 (1979): 513–543.
- Montague, R. *Formal Philosophy*, edited by R. Thomason. New Haven, CT: Yale University Press, 1974.
- Quine, W. V. O. *Word and Object*. Cambridge, MA: MIT Press, 1960.
- Richard, M. *Propositional Attitudes*. Cambridge, U.K.: Cambridge University Press, 1990.
- Salmon, N., and S. Soames, eds. *Propositions and Attitudes*. Oxford: Oxford University Press, 1988.
- Schiffer S. “A Paradox of Meaning.” *Noûs* 28 (1994): 279–324.
- Stalnaker, R. *Inquiry*. Cambridge, MA: MIT Press, 1984.

**Mark E. Richard (1996)**

## PROPOSITIONS [ADDENDUM]

Despite the rearguard efforts of Robert Stalnaker and Max Cresswell, by the late 1990s it became widely acknowledged that sets of possible worlds are too coarse-grained to serve as propositions. It is safe to say that among those philosophers who believe in propositions, most think of them as sententially structured entities, composed out of the contents of the words and phrases in the sentences that express them. Fregeans hold that these contents are Fregean senses; Russellians hold that they are objects, properties, and relations.

Yet the 1990s also saw new challenges and approaches to structured propositions. George Bealer and Michael Jubien have independently argued (i) that it is counterintuitive to hold that we believe and assert structured complexes, and (ii) that theories of structured propositions are subject to the same problem that Paul Benacerraf raised for set-theoretic reductions of arithmetic. On one such reduction, the number 2 is identified with the set  $\{\{\emptyset\}\}$ ; on another, 2 is identified with  $\{\emptyset, \{\emptyset\}\}$ , where  $\emptyset$  is the null set. Benacerraf's problem is that there are no principled reasons for preferring one or the other reduction, or any of the infinitely many equally good alternatives, and so none of these reductions can be correct. For similar reasons, the proposition that Jones loves Smith cannot be identified with the ordered set  $\langle \text{love}, \langle \text{Jones}, \text{Smith} \rangle \rangle$ , or with  $\langle \text{Jones}, \langle \text{love}, \text{Smith} \rangle \rangle$ , or ....

These and other problems led Bealer to reject all reductions of propositions to structured objects and to hold that propositions are unstructured and irreducible. They led Jubien to reject propositions altogether in favor of a Russellian multiple-relation theory of judgment, which dispenses with propositions by analyzing "believes" and other attitude verbs as many-place predicates that relate subjects to objects, properties, and relations instead of to whole propositions. However, (i) is debatable, and (ii) can be avoided if one can provide a rationale for preferring one system of reduction. For example, Jeffrey King holds (roughly) that a structured proposition is obtained by replacing the words of a sentence with their contents while retaining the syntactic relations in the logical form of the sentence. This solves Benacerraf's problem because the structure in propositions is identified with the syntactic structure in the logical form. The connection with syntax provides a principled reason for identifying propositions with the structured objects proposed by King.

Another approach to structured propositions is due to Jon Barwise and John Etchemendy, who use what they call "Austinian propositions," named after the Oxford philosopher J. L. Austin, in their solution to the liar paradox. An Austinian proposition is like a structured Russellian proposition except that it contains a contextually determined situation that the proposition is about. So while the Russellian proposition that Claire is playing cards is true just in case Claire is playing cards, the Austinian proposition that Claire is playing cards is true just in case Claire is playing cards in the contextually determined situation. For every situation  $s$ , there is a liar proposition  $f$  about  $s$  that claims that  $f$  is false in  $s$ . In Barwise and Etchemendy's formal development, it turns out that every such  $f$  is simply false. However, for every  $s$ , there is an expanded situation  $s'$ , and there is a true proposition  $p$  about  $s'$  that claims that  $f$  is false in  $s'$ . The intuition that the liar proposition  $f$  is both true and false arises out of a failure to keep separate the distinct Austinian propositions  $f$  and  $p$ .

A general challenge to propositions has come from Donald Davidson, who has used the so-called slingshot argument to collapse all facts into a single Great Fact, effectively robbing facts of their philosophical utility. Davidson argues that if we give up on facts, we should also give up on entities that represent facts, such as propositions. The slingshot argument can also be used directly against propositions to show that all true propositions collapse into a single Great Proposition. But as Stephen Neale has shown, the slingshot argument can be avoided as long as one holds that sentential operators like "the fact that ... is identical to the fact that ..." and "the proposition that ... is identical to the proposition that ..." satisfy certain logical constraints on inference rules involving definite descriptions. This constraint can easily be satisfied if one adopts a Russellian analysis of definite descriptions (which construes "The  $f$  is  $g$ " as "There is exactly one  $f$ , and it is  $g$ ").

**See also** Meaning; Propositional Attitudes: Issues in Semantics.

### **Bibliography**

- Barwise, Jon, and John Etchemendy. *The Liar*. New York: Oxford University Press, 1987.
- Bealer, George. "Propositions." *Mind* 107 (425) (1998): 1–32.
- Benacerraf, Paul. "What Numbers Could Not Be." *Philosophical Review* 74 (1965): 47–73.
- Cresswell, Max. *Structured Meanings*. Cambridge, MA: MIT Press, 1985.
- Davidson, Donald. "The Structure and Content of Truth." *Journal of Philosophy* 87 (1990): 279–328.

Jubien, Michael. "Propositions and the Objects of Thought." *Philosophical Studies* 104 (2001): 47–62.

King, Jeffrey. "Structured Propositions and Complex Predicates." *Noûs* 29 (1995): 516–535.

Neale, Stephen. *Facing Facts*. New York: Oxford University Press, 2001.

Stalnaker, Robert. *Inquiry*. Cambridge, MA: MIT Press, 1984.

*Peter W. Hanks (2005)*

## PROTAGORAS OF ABDERA

Protagoras of Abdera in Thrace, most famous of the Sophists, was born not later than 490 BCE and probably died soon after 421 BCE. According to Plato, he was the first to declare himself a professional Sophist. He went from city to city in the Greek world, offering instruction in return for money, and he undertook above all to train young men in the art of politics. He was well known in Athens, where he enjoyed the friendship of Pericles—he produced a theoretical basis for Periclean democracy and was asked by Pericles to draft the constitution for the new colony of Thurii in 443 BCE. He made contributions to grammatical and rhetorical theory, and his views on religion provoked charges of impiety against him in the courts, which led to his exile from Athens at the end of his life and to the public burning of at least one of his books.

His writings were numerous and included "On Truth," "On the Gods," and "Antilogic" (or "Antilogies"). Later writers probably took their information about him mainly from the accounts of Plato, Aristotle, and Sextus Empiricus, but one of his works was read by Porphyry in the third century CE, and in the Hellenistic period he was regarded as sufficiently important for his statue to be set up, together with those of Plato, Aristotle, and other thinkers, in the Serapeum at Memphis in Egypt.

Since the time of Plato, Protagoras's main doctrines have been regarded as possessing considerable philosophical interest, even by those who deny philosophical importance to the Sophists in general; but very divergent interpretations have been propounded. With no surviving works and virtually no fragments, interpretation must depend upon the assessment of the evidence of Plato, Aristotle, and Sextus Empiricus. In what follows, the view is taken that Plato in the *Theaetetus* correctly states the basic position of Protagoras and then proceeds to distinguish certain possible developments of this position not held by Protagoras. The basic position was independently understood in the same way by both Aristotle and Sextus Empiricus, each of whose information was

not simply derived from the *Theaetetus*. This would be denied by some scholars.

### EPISTEMOLOGY

The starting point must be the famous contention that "man is the measure of all things, of things that are that [or 'how'] they are and of things that are not that [or 'how'] they are not." Theodor Gomperz maintained that "man" is to be understood collectively in the sense of "mankind as a whole" or "the human race." But against this, the evidence of the *Theaetetus* 152A–B seems to show conclusively that it is individual men that Protagoras had in mind in the first instance, although, as will be seen, his theory is capable of easy extension to groups of men, and he probably made this extension himself.

According to Plato's example in the *Theaetetus*, when the same wind appears cold to one person and warm to another person, then the wind is warm to the person to whom it appears warm and is cold to the person to whom it seems cold. It follows that all perceptions are true and the ordinary view is mistaken, according to which, in cases of conflict, one person is right and the other person is wrong about the quality of the wind or of anything else. This clearly was the position held by Protagoras, but it is not clear exactly how he came to this view. It is often held that his position is a kind of subjective idealism similar to that of Bishop Berkeley, according to which qualities in a thing are for the person to whom they seem, so long as they seem to him, but have no existence independent of their seeming.

Against this view, Sextus Empiricus is explicit: All qualities perceived by different persons are actually present in matter. Sextus's introduction of matter may well be anachronistic, but his account suggests an alternative view, accepted by F. M. Cornford among others, according to which opposite qualities are copresent in objects, and in cases of conflict of perceptions between two persons, what happens is that we have a sort of selective perception—one person perceives one quality and the other its opposite, both qualities being present in the situation, waiting to be perceived, as it were, independently of any actual perceiving by a subject. This view seems to have the support of Aristotle, who always treats Protagoras's doctrine as involving the denial of the principle of contradiction, and the view coincides with incidental pointers in Plato's account ("the same wind"—152B; "perception, then, is always of something that is"—152C). It is true that in the "secret doctrine" attributed to Protagoras by Plato (152Cff.) the independent status of sense objects is undermined, but the fact that this is presented as a secret

doctrine is surely conclusive evidence that it was a doctrine not publicly associated with Protagoras.

The “man-measure” doctrine is presented by Plato in the first instance as a doctrine about perception of sensible qualities. But it is clear that Plato supposed that for Protagoras it also applied to moral and aesthetic qualities such as “just” and “beautiful.” It is especially in these cases that the extension of the doctrine to groups of people was made by Protagoras—“whatever seems just to a city is just for that city so long as it seems so.” Probably Protagoras did not extend his doctrine to apply to all judgments; this was done immediately by his opponents in the famous *peritrope*, or “turning of the tables”: Let us suppose that whatever seems true to any person is true for the person to whom it seems so. If this is the doctrine of Protagoras, then Protagoras will hold that those who hold that Protagoras’s theory is false are holding the truth (*Theaetetus* 171A). But Plato points out that if Protagoras could pop his head up through the ground, he would surely have an answer to this objection.

At the very least, Protagoras was clear about one point. In the case of conflict about perceived qualities all perceptions are true. But some perceptions are better than others, for example, the perceptions normally found in a healthy man as distinct from those found in a man who is ill. It is the function of a doctor, Protagoras held, to change a man who is ill so that his perceptions become those of a man who is well. Likewise, in moral, political, and aesthetic conflicts it is the function of the Sophist as a teacher to work a change so that better views about what is “just” and “beautiful” will seem true to the “patient”—better, that is, than those that previously seemed true to him. All the “patient’s” views are equally true, but some are better than others.

There is nothing to suggest that by “better” Protagoras meant what will seem better. Quite the contrary. Better views are views that have better consequences, and consequences which are better are so as a matter of fact, independently of whether a person thinks them better or not. In other words, Protagoras here made an exception to his man-measure doctrine. There is every reason to suppose that he would have excepted the class of judgments about the consequences of judgments from his principle. Indeed, there is no actual evidence in any ancient author that Protagoras himself ever applied his doctrine to statements other than those about perceived qualities and moral and aesthetic qualities treated on the same plane as visually perceived qualities. What probably happened was that he propounded his doctrine in certain general statements such as “whatever seems to anyone is

so for that person,” without adding the qualifications that he really intended; thus he gave a handle to his enemies, which enabled them to apply the *peritrope* and similar objections.

The above account rests primarily upon Plato’s *Theaetetus*. To it may be added evidence from other sources. According to Diogenes Laërtius, Protagoras was the first to propound the theory that there are two *logoi*, or accounts, to be given about everything. This has sometimes been treated as simply the now familiar rhetorical doctrine that “there are two sides to every question.” But this theory was used as a method of argument, and it should probably be related to the man-measure doctrine and to what Plato called “Antilogic,” the probable title of one of Protagoras’s treatises. In conflicts about perceived qualities, and also moral and aesthetic qualities, there might seem room for an infinite variety of “seemings,” but if we take any one as a starting point, for instance, that the wind seems warm, all other seemings may be expressed as the negative of this, namely “not-warm.” This was clearly the way in which Plato tended to regard phenomena—as did the antilogicians, too—namely, as always being both “warm” and “not-warm.” In this view, Plato was probably following Protagoras. It is possible that Protagoras associated with the two-*logoi* principle the prescription attributed to him by Aristotle “to make the lesser [or ‘the weaker’] argument the stronger.” This may have been what the Sophist was expected to do when altering a man’s opinions for the better.

## SOCIAL THEORY

In Plato’s dialogue *Protagoras* we are given a coordinated theory of the Sophist in relation to society and of a possible theoretical basis for a Periclean-style democracy. All is completely consistent with the positions attributed to Protagoras in the *Theaetetus*. When Protagoras professes to make men good citizens, Socrates objects that while the Athenians call in experts to advise on technical matters, they regard all citizens as capable of advising them on matters relating to the city. This seems to imply that Athenian democracy leaves no place for expert instruction in citizenship. Protagoras replies with a myth followed by a nonmythical exposition that while all men share in the qualities that make good citizens, they do not do so by nature but acquire these qualities by instruction and by practice. These qualities are beliefs and opinions about what is just and right. In a sense, the whole community teaches its members about these matters, and so all are rightly consulted about political matters. But the expert teacher, such as the Sophist Protagoras, can



improve opinions on such matters, whether it be in the case of an individual or in the case of a whole community.

## OTHER VIEWS

Protagoras's doctrines ranged beyond the topics discussed above to cover physical and mathematical problems as well, but it is no longer possible to state his actual teachings on these problems. He seems to have held that a tangent touches a circle not only at one point, but at more than one, clearly arguing from visual experience of drawn lines. Parmenides had rejected the world of seeming in favor of his world of being; Protagoras took the opposite path and attempted to expound a world in which all appearances were true and where there was nothing outside or beyond what appeared. This involved the copresence of opposed and contradictory qualities at many points. Protagoras was prepared to accept and explain this copresence through his "man-measure" principle, either on the basis of a theory of subjective idealism or, more probably, on the basis of a conception of a phenomenal world actually composed of opposites (a conception typical of the pre-Socratics). This conception seemed to Plato to be substantially correct for the phenomenal world, hence his great interest in Protagoras. But Plato felt that this view made it impossible to give any account or explanation of phenomena, and to be able to give an explanation seemed to him essential.

Diogenes Laërtius says that for Protagoras the soul is nothing apart from its perceptions. This suggests a phenomenalistic view of the soul as well as of everything else. Diogenes' account may be correct, although doubts have been cast upon it. If it is correct, however, it probably was not intended to imply any doctrine like the modern theory of neutral monism, but simply to deny the existence of any "submerged," or nonphenomenal, element in the soul.

*See also* Ethical Relativism; Sophists.

## Bibliography

### TEXTS AND TRANSLATIONS

- Gagarin, Michael, and Paul Woodruff. *Early Greek Political Thought from Homer to the Sophists*. Cambridge, U.K.: Cambridge University Press, 1995.
- Sprague, R.K., ed. *The Older Sophists; a Complete Translation by Several Hands of the Fragments in die Fragmente der Vorsokratiker*, ed. by Diels-Kranz. Columbia: University of South Carolina Press, 1972.
- Fragments and testimonia of Protagoras' work may be found in H. Diels and W. Kranz, eds., *Fragmente der Vorsokratiker*,

10th ed. (Berlin, 1961), Vol. II. A. Capizzi, *Protagora* (Florence, 1955), is probably the best full discussion, and it includes some material not in Diels and Kranz.

### STUDIES

- Bett, R. "The Sophists and Relativism." *Phronesis* 34 (1989): 139–169.
- Classen, J. "Protagoras' *Aletheia*." In *The Criterion of Truth, Essays Written in Honour of George Kerferd*, edited by P. Huby and G. Neal. Liverpool: Liverpool University Press, 1989.
- Guthrie, W. K. C. *The Sophists*. Cambridge, U.K.: Cambridge University Press, 1971.
- Kerferd, G. B. "Plato's Account of the Relativism of Protagoras." *Durham University Journal* 42 (1949): 20–26.
- Kerferd, G. B. "Protagoras' Doctrine of Justice and Virtue in the 'Protagoras' of Plato." *Journal of Hellenic Studies* 73 (1953): 42–45.
- Kerferd, G. B. *The Sophistic Movement*. Cambridge, U.K.: Cambridge University Press, 1981.
- Schiappa, E. *Protagoras and Logos: A Study in Greek Philosophy and Rhetoric*. Columbia, SC: University of South Carolina Press, 1991.
- Vlastos, G. "Introduction." In *Plato's Protagoras*, translated by M. Ostwald. Indianapolis/New York: Bobbs-Merrill, 1956.

**G. B. Kerferd (1967)**

*Bibliography updated by Paul Woodruff (2005)*

## PROTOCOL SENTENCES

*See Basic Statements*

## PROUDHON, PIERRE-JOSEPH (1809–1865)

Pierre-Joseph Proudhon has been called the father of anarchism, a title that is accurate insofar as organized anarchist movements throughout the world can be traced to his teachings and to the actions of his disciples. Proudhon was also the first writer deliberately to accept the title of anarchist, which he did in 1840. Before his time the term had been used to denote one who seeks to promote social disorder; Proudhon argued that it could be used with more justice to describe one who seeks social order without authoritarian government. "As man seeks justice in equality, so society seeks order in anarchy," he said. "Anarchy—the absence of a master, of a sovereign—such is the form of government to which we are every day approximating." Such doctrines were not entirely original; the English writer William Godwin had expounded them fifty years earlier without describing them as "anar-

chist,” but Proudhon appears to have been uninfluenced by Godwin and to have reached his conclusions independently.

Proudhon prided himself on being a man of the people. He was born in Besançon, capital of Franche-Comté, of Jura peasant stock. His childhood was hard, and after a brief period at the college in Besançon, he received his education largely through his work as a printer; he taught himself Greek and Hebrew and developed a prose style that eventually won the admiration of Charles-Pierre Baudelaire, Gustave Flaubert, and Victor Hugo. The turning point in Proudhon’s career came when he was awarded a scholarship by the Besançon Academy in 1838. This took him to Paris and gave him the leisure to formulate his ideas and to write his first important book, *Qu’est-ce que la propriété?* (*What Is Property?*, Paris, 1840). This book, hailed by Karl Marx as “the first decisive, vigorous and scientific examination” of the institution of property, gained notoriety because in one passage Proudhon defined property as “theft.” The author’s love of telling phrases distorted the nature of his argument, for *Qu’est-ce que la propriété?* was in fact an investigation of abuses that had entered into the institution of property rather than a condemnation of property itself. The arguments that Proudhon put forward in this early book, on the nature of property and the faults of government, are those which he elaborated and gave a deeper philosophical backing in his later works.

Proudhon attacked the existence of private property that allows the exploitation of the labor of others, such as the owning of land by those who do not work it; he had only approval for the “possession” that allows a worker to dispose of what his hands make. “The right to products is exclusive—*jus in re*; the right to means is common—*jus ad rem*.” This is so because the means of production, the heritage of techniques and inventions, have been built up by human cooperation, and no man has a right to use them exclusively for his own benefit. However, for the sake of independence, Proudhon granted the need for each man to control the land or tools he can use. In this early book he still thought in terms of a peasant-and-handcraft society.

Proudhon attacked unreformed property because it negates equality, but he rejected the communist theories of his time (principally those of the French utopian socialists) because they denied independence. Here Proudhon came to the political aspect of his argument—both unreformed property and communism are dependent on forms of authority to maintain themselves. But how far is authority justified? Proudhon contended that it

arises from the tendency of social animals and primitive man to seek leaders. As reason develops, criticism, protest, and rebellion arise. Emergent political science finds the laws by which society functions in the nature of things, not in the whims of rulers. At this point anarchy, administration without government, becomes possible. Proudhon, at this stage under the influence of Hegelian ideas imperfectly absorbed from French reviews, created a triad. The thesis is property, which destroys equality; the antithesis is communism, which denies independence; the synthesis is anarchy or liberty, which is embodied in a society of producers bound together by a network of free contracts. In the widening recognition of mutual interests, government becomes unnecessary.

During the 1840s Proudhon served for several years as office manager for a water transport firm in Lyons, work that allowed him to travel frequently to Paris. In these two settings his theory of mutualism—the form of anarchism particularly associated with him—developed. Political radicalism flourished in mid-nineteenth-century Lyons, and Proudhon encountered there the disciples of Étienne Cabet, Charles Fourier, Pierre Leroux, and other socialist prophets. He developed the idea of a worldwide working-class organization on an economic basis rather than a political one. This led him to place faith in various forms of mutual credit systems that might eventually make governmental administration unnecessary; he envisaged such associations as becoming worldwide. In Paris, Proudhon associated with some of the leading European revolutionary theorists, including Marx, Mikhail Bakunin, and Alexander Ivanovich Herzen. However, his personal and theoretical incompatibility with Marx soon became evident; the historic conflict between libertarian and authoritarian views of socialism began with the split between Marx and Proudhon, which dates from Marx’s attack in *La misère de la philosophie* (Paris, 1847) on Proudhon’s *Système des contradictions économiques* (2 vols., Paris, 1846). Bakunin and Herzen, on the other hand, eventually became Proudhon’s most important disciples.

During the 1840s Proudhon, an eclectic thinker, took what he found valid from the writings of G. W. F. Hegel, Ludwig Feuerbach, Immanuel Kant, and other German philosophers, as well as from Auguste Comte and the French utopians. He evolved a philosophy that left out the third term of the Hegelian triad, and accepted contradiction as an enduring force tending toward a dynamic equilibrium—the desirable condition of existence. He denied all absolutes, all utopian aspirations to permanent solutions, and, in his *Philosophie du progrès* (Paris, 1853) saw

progress as “the affirmation of universal movement and in consequence the negation of all immutable forms and formulae, of all doctrines of eternity, permanence, or impeccability, and of every subject, or object, spiritual or transcendental, that does not change.” He was, deliberately and avowedly, an antisystematic philosopher.

Proudhon assumed the standpoint of a critical independent, and as such he became the most outspoken journalist of the period, giving qualified support to the French revolution of 1848. His *Le représentant du peuple* (1848) was the first anarchist newspaper published with any regularity; harried by suppressions and fines, it survived under various names for more than two years. Proudhon was elected in June 1848 to the Constituent Assembly, where he maintained an intransigent minority position. He also planned a people’s bank, based on his mutualist ideas, which never materialized because he was imprisoned for attacks in his paper on Louis Napoleon, then president of the Republic.

Proudhon’s three years of imprisonment were light: He was allowed occasional days out on parole, on one of which he married Euphrasie Piégard, and he wrote two of his most important books, *Les confessions d’un révolutionnaire* (Paris, 1850), an analysis of the events of 1848 that states the aim of anarchist revolutionism as “no more government of man by man, by means of the accumulation of capital,” and *Idée générale de la révolution au XIX<sup>e</sup> siècle* (*General Idea of the Revolution in the Nineteenth Century*, Paris, 1851). The latter book comes nearer than anything else Proudhon wrote to presenting his view of the ideal libertarian society, based on contract instead of laws, with authority decentralized in communes and industrial associations, with frontiers abolished and flexible federation replacing the centralized national state.

During the early years of the Second Empire, Proudhon was subjected to constant police persecution, and in 1858 he was again sentenced to three years’ imprisonment for an offense against the press laws. He fled to Belgium, where, although pardoned in 1860, he lived until 1862. During his final years in Paris, he gained a considerable mutualist following among French workingmen, and before he died early in 1865, he learned that his followers had taken a leading part in the meetings that led to the founding of the International Workingmen’s Association.

During his final years Proudhon wrote a number of books that elaborated important aspects of his doctrines. *Du Principe fédératif* (Paris, 1863) summarized his criticism of nationalism and developed his ideas of communal organization leading gradually to world federation.

*De la Justice dans la révolution et dans l’église* (3 vols., Paris, 1858) opposed his own theory of an immanent justice to transcendentalist ideas of justice. *De la Capacité politique des classes ouvrières* (published posthumously in Paris, 1865) developed Proudhon’s view of the power of the working class to achieve its own liberation by economic means.

Later anarchism and syndicalism were largely influenced by Proudhon’s doctrines, as was the populist movement in Russia. As the Russian anarchist Mikhail Bakunin said, “Proudhon was the master of us all.”

**See also** Bakunin, Mikhail Aleksandrovich; Comte, Auguste; Feuerbach, Ludwig Andreas; Fourier, François Marie Charles; Hegel, Georg Wilhelm Friedrich; Herzen, Aleksandr Ivanovich; Kant, Immanuel; Marx, Karl.

### Bibliography

#### WORKS BY PROUDHON

*Oeuvres complètes de P. J. Proudhon*, 26 vols. Paris, 1867–1870.

*Correspondance*, 14 vols. Paris, 1874–1875.

*What Is Property?* Translated by Benjamin Tucker. Princeton, MA: B.R. Tucker, 1876.

*System of Economical Contradictions*. Translated by Benjamin Tucker. Boston, 1888.

*Oeuvres complètes de P. J. Proudhon*. Edited by C. C. A. Bouglé and Henri Moysset, 11 vols. Paris, 1920–1939. Never completed.

*General Idea of the Revolution in the Nineteenth Century*. Translated by John Beverley Robinson. London: Freedom Press, 1923.

#### WORKS ON PROUDHON

Brogan, D. W. *Proudhon*. London, 1934.

Dolléans, Édouard. *Proudhon*. Paris: Gallimard, 1948.

Hoffman, R. L. *Revolutionary Justice: The Social and Political Thought of P.-J. Proudhon*. Urbana: University of Illinois Press, 1972.

Hyams, Robert S. *Pierre-Joseph Proudhon: His Revolutionary Life, Mind, and Works*. London: Murray, 1979.

Lubac, Henri de. *Proudhon et le christianisme*. Paris, 1945. Translated by R. E. Scantlebury as *The Un-Marxian Socialist: A Study of Proudhon*. London: Sheed and Ward, 1948.

Noland, Aaron. “Proudhon and Rousseau.” *Journal of the History of Ideas* 28 (1967): 33–54.

Prion, Gaëtan. *Proudhon et syndicalisme révolutionnaire*. Paris, 1910.

Ritter, Alan Irving. *The Political Thought of Pierre-Joseph Proudhon*. Princeton, NJ: Princeton University Press, 1969.

Sainte-Beuve, Charles A. *P.-J. Proudhon*. Paris, 1872.

Vincent, K. Steven. *Pierre-Joseph Proudhon and the Rise of French Republican Socialism*. New York: Oxford University Press, 1984.

Woodcock, George. *Pierre-Joseph Proudhon*. London: Routledge and Kegan Paul, 1956.

*George Woodcock (1967)*  
*Bibliography updated by Philip Reed (2005)*

## PROUST, MARCEL

(1871–1922)

The French author Marcel Proust was born and educated in Paris. He lived there all his life, leaving only for short holidays or artistic pilgrimages, most of which were to the great cathedral cities of France. His father, a professor of medicine, was Catholic; his mother, whom he adored, was Jewish. Both traditions, as well as his consuming interest in French history and culture, played important roles in his life and art, although he was neither religiously orthodox nor politically chauvinistic. He undertook a considerable and seemingly futile search for a vocation and did some writing, most of which was discarded drafts of his future novel. Suffering terribly from asthma and from certain guilts about his homosexuality, but with economic as well as spiritual means sufficient to indulge and transmute these ills, Proust ensconced himself in his famous cork-lined room to write his masterpiece, *A la recherche du temps perdu*.

### PHILOSOPHICAL THEMES

Although Proust compared a work of art in which there are theories to an object on which the price is marked, *A la recherche* is, nonetheless, a philosophical novel. There are two major philosophical themes woven into the novel: that reality is composed of artistic essences and that the search for essences ends in their dissolution. Proust stated only the first theme; the second, however, is implied by much of the action of the novel.

In the last volume of the novel, *Le temps retrouvé*, Proust, as narrator and participant, stated his theory of artistic essences as reality; this theory, because of its role in the context of the whole novel, must be understood as an integral part of it, along with the characterization, dialogue, and plot. According to Proust's theory, we live in a world of people, places, and things, all of which are organized spatially or temporally, in the ordinary sense of space and time, and which impinge on us. Most of us merely react to these phenomena. The true artist, however, like the scientists, attempts to find the laws that govern these phenomena. Whereas the scientist proceeds by his intellect, the artist cannot, for his laws are to be discovered only by intuition. The artist's intellect supple-

ments, but it cannot supplant, intuition. Intuition is that state of mind in which the artist—rooted in past experiences, nourished by suffering, and graced by an involuntary memory of a past sensation joined with a similar present one—extracts the qualitative similarity or essence from these sensations in order to embody that essence in a metaphor which, like the essence, is not subject to the ravages of time. Thus, these essences are the only true reality, and their artistic expression the only true judgment on reality.

Proust, it is important to realize, did not deny the existence of temporal or spatial relations, but he rejected them as unreal. Hence, he must understand by *reality* something quite distinct from *existence*: *reality* for him functioned as an honorific term denoting that which is salvageable from the past and which transcends the present—that, therefore, which is ultimate in the precise sense of being out of time. *Reality*, in effect, denotes the essences extracted by intuition from what exists in relation to what existed.

It has been claimed that Proust's conceptions of time and intuition are Bergsonian. It seems, however, that there are important differences. According to Henri Bergson, time is essentially duration (*durée*). The concepts of the past, present, and future cannot apply to time because they spatialize it. Duration can only be experienced, not thought of or talked about; it is the indivisible, ultimate fact of process in the world, and intuition is the experience of duration, a direct acquaintance with it. For Proust, however, time is not duration; it consists of chronological relations among events. Nor is time ultimate; only the timeless essences are that. Finally, intuition for Proust is an extraction from, not an immersion in, time.

Nor is Proust's theory Platonic, as has sometimes been suggested. Plato's timeless essences are perfect and have their being absolutely independently of the spatial and temporal particulars of this world; the Proustian essences are at most more or less imperfect copies of the truly real forms.

Besides this aesthetic-ontological theme, which Proust integrated magnificently in the novel, there is the nether theme of the dissolution of essences in the very search for them. Although he never stated this theme, much of the novel embodies it. The treatment of love is probably the best single example. Through the narration of many different love relationships, commonly regarded as a major achievement of the novel, Proust dramatized that love has no essence, only an inexhaustible set of properties, none of which is necessary or sufficient. Here

intellect supplants rather than supplements intuition. Proust's observations, analyses, and generalizations harvest a vast multiplicity of criteria that govern our understanding and concept of love. In effect, Proust showed through his characterization, monologue and dialogue, as well as through the plot, that the range of the experience of love renders impossible any traditional essentialist definition of it. To have discovered, explored, and artistically wrought this important truth about our conceptual life and to have shown it a full generation before philosophers stated it is not the least of Proust's accomplishments in his great novel.

**See also** Appearance and Reality; Bergson, Henri; Intuition; Plato; Platonism and the Platonic Tradition.

### Bibliography

#### WORKS BY PROUST

- Portraits de peintres*. Paris: Heugel, 1896.
- Les plaisirs et les jours*. Paris, 1896. Translated by Louise Varese and others as *Pleasures and Days*, with an introduction by F. W. Dupee. Garden City, NY: Doubleday, 1957.
- La bible d'Amiens*. Paris, 1904. A translation, with preface and notes, of John Ruskin's *Bible of Amiens*.
- Sésame et lys*. Paris, 1906. A translation, with preface ("Journées de lecture") and notes, of Ruskin's *Sesame and Lilies*.
- A la recherche du temps perdu* (published by the *Nouvelle Revue Française*). 7 parts, 13 vols. Paris, 1919–1927. One volume had been published in 1913, at Proust's expense. Translated by C. K. Scott Moncrieff (the first 6 parts) and Stephen Hudson (last part only) as *Remembrance of Things Past*. New York: Random House, 1922–1931. Uniform edition, 12 vols. London and New York, 1941. Pléiade edition, 3 vols., Paris, 1954.
- Chroniques*. Paris: Éditions de la Nouvelle revue française, 1927.
- Jean Santeuil*. With introduction by André Maurois. Paris, 1954. Translated by Gerard Hopkins. London: Weidenfeld and Nicolson, 1955; New York: Simon and Schuster, 1956.
- Contre Sainte-Beuve*. Paris: Gallimard, 1954. Translated by Sylvia Townsend Warner as *By Way of Sainte-Beuve*. London: Chatto and Windus, 1958.
- #### WORKS ON PROUST
- Brée, Germaine. *Marcel Proust and Deliverance from Time*. New Brunswick, NJ: Rutgers University Press, 1955.
- Cocking, J. M. *Proust*. London: Bowes and Bowes, 1956.
- Delattre, Floris. *Bergson et Proust*. Paris, 1948. Vol. I of *Les études Bergsoniennes*.
- Green, F. C. *The Mind of Proust*. Cambridge, U.K.: Cambridge University Press, 1949.
- Maurois, André. *The Quest for Proust*. London, 1950.

Weitz, Morris. *Philosophy in Literature: Shakespeare, Voltaire, Tolstoy and Proust*. Detroit: Wayne State University Press, 1963.

Morris Weitz (1967)

## PROVABILITY LOGIC

Even though "provability logic" did not come into its own until the early seventies, it has its roots in two older fields: metamathematics and modal logic. In metamathematics, we study what theories can say about themselves. The first—and most outstanding—results are Kurt Gödel's two incompleteness theorems.

If we take a sufficiently strong formal theory  $T$ —say, Peano arithmetic—we can use Gödel numbering to construct in a natural way a predicate  $Prov(x)$  in the language of  $T$  that expresses " $x$  is the Gödel number of a sentence which is provable in  $T$ ." About  $T$  we already know that it satisfies *modus ponens*:

If it is provable that  $A$  implies  $B$ , then, if  $A$  is provable,  $B$  is provable as well.

Now it turns out that, using Gödel numbering and the predicate  $Prov$ , we can express *modus ponens* in the language of  $T$ , and show that in  $T$  we can actually prove this formalized version of *modus ponens*:

$$Prov(\ulcorner A \rightarrow B \urcorner) \rightarrow (Prov(\ulcorner A \urcorner) \rightarrow Prov(\ulcorner B \urcorner)).$$

When we rephrase both the normal and the formalized version of *modus ponens* using the modal operator  $\Box$ , reading  $\Box A$  as " $A$  is provable in  $T$ ," we get the modal rule

$$(1) \quad \frac{A \rightarrow B \quad A}{B}$$

and the modal axiom

$$(2) \quad \Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B).$$

Indeed both the rule and the axiom are well known from the basic modal logic  $K$ .

Similarly, we can show that if there is a proof of the sentence  $A$  in  $T$ , then  $T$  itself can check this proof, so  $T$  proves  $Prov(\ulcorner A \urcorner)$ —we shall call this principle *Prov-completeness*. Again, though in a less straightforward way than in the case of *modus ponens*, we can formalize the principle itself and see that  $T$  actually proves:

$$Prov(\ulcorner A \urcorner) \rightarrow Prov(\ulcorner Prov(\ulcorner A \urcorner) \urcorner).$$

When we rephrase the principle of *Prov*-completeness and its formalization in modal logical terms, we get the modal rule that is usually called necessitation:

$$(3) \quad \frac{A}{\Box A},$$

and the modal axiom

$$(4) \quad \Box A \rightarrow \Box \Box A,$$

which is the transitivity axiom 4 well known from modal systems such as K4 and S4.

Finally, one might wonder whether *T* proves the intuitively valid principle that “all provable sentences are true,” that is, whether *T* proves  $\text{Prov}(\ulcorner A \urcorner) \rightarrow A$ . Unexpectedly, this turns out not to be the case at all. Löb proved in 1953, using Gödel’s technique of diagonalization, that *T* proves  $\text{Prov}(\ulcorner A \urcorner) \rightarrow A$  only in the trivial case that *T* already proves *A* itself!

Löb’s theorem has a formalization that can also be proved in *T*. Writing both the theorem and its formalization in modal terms, we get the modal rule

$$(5) \quad \frac{\Box A \rightarrow A}{A},$$

and the modal axiom

$$(6) \quad \Box(\Box A \rightarrow A) \rightarrow \Box A,$$

usually called *W* (for well-founded) by modal logicians.

Now we can define provability logic, which goes by various names in the literature—*PRL*, *GL* (for Gödel/Löb), *L* (for Löb), and, in modal logic texts, *KW4*. It is generated by all the modal formulas that have the form of a tautology of propositional logic, plus the rules (1),(3),(5) and axioms (2),(4),(6) given above. One can prove that rule (5) and axiom (4) already follow from the rest, so that *PRL* is equivalently given by the well-known system *K* plus the axiom  $\Box(\Box A \rightarrow A) \rightarrow \Box A$ .

The main “modal” theorem about *PRL*—but one with great arithmetical significance—is the “fixed point theorem,” which D. de Jongh and G. Sambin independently proved in 1975. The theorem says essentially that “self-reference is not really necessary.” Suppose that all occurrences of the propositional variable *p* in a given formula *A* are under the scope of  $\Box$ -es, for example,  $A(p) = \neg \Box p$  or  $A(p) = \Box(p \rightarrow q)$ . Then there is a formula *B* in which *p* does not appear, such that all propositional variables that occur in *B* already appear in  $A(p)$ , and such that  $PRL \vdash B \leftrightarrow A(B)$ . This *B* is called a fixed point of  $A(p)$ .

Moreover, the fixed point is unique, or more accurately, if there is another formula *B'* such that  $PRL \vdash B' \leftrightarrow A(B')$ , then we must have  $PRL \vdash B \leftrightarrow B'$ . Most proofs of the fixed point theorem in the literature give an algorithm by which one can compute the fixed point.

For example, suppose that  $A(p) = \neg \Box p$ . Then the fixed point produced by the algorithm is  $\neg \Box \perp$ , and indeed we have  $PRL \vdash \neg \Box \perp \leftrightarrow \neg \Box(\neg \Box \perp)$ . If we read this arithmetically, the direction from left to right is just the formalized version of Gödel’s second incompleteness theorem. Thus, if *T* does not prove a contradiction, then it is *not* provable in *T* that *T* does not prove a contradiction.

The landmark result in provability logic is Solovay’s “arithmetical completeness theorem” of 1976. This theorem says essentially that the modal logic *PRL* captures *everything* that Peano arithmetic can say in modal terms about its own provability predicate. Before formulating Solovay’s theorem more precisely, we turn to the semantics of *PRL*.

Provability logic has a suitable Kripke semantics, just like many other modal logics. Unaware of the arithmetical relevance of *PRL*, Krister Segerberg proved in 1971 that it is sound and complete with respect to finite irreflexive transitive frames, and even with respect to finite trees. This completeness theorem immediately gives a decision procedure to decide for any modal formula *A* whether *A* follows from *PRL* or not. Looking at the procedure a bit more precisely, it can be shown that *PRL* is “very decidable”: Like the well-known modal logics *K*, *T*, and *S4*, it is decidable in PSPACE. This means that there is a Turing machine that, given a formula *A* as input, answers whether *A* follows from *PRL*; the size of the memory that the Turing machine needs for its computations is only polynomial in the length of *A*.

The modal completeness theorem was an important first step in Solovay’s proof of the arithmetical completeness of *PRL*. Suppose that *PRL* does not prove the modal formula *A*. Then there is a finite tree such that *A* is false at the root of that tree. Now Solovay devised an ingenious way to describe the tree in the language of Peano arithmetic. Thus he found a translation *f* from modal formulas to sentences of arithmetic, such that Peano arithmetic does not prove  $f(A)$ . Such a *translation f* respects the logical connectives (so, e.g.,  $f(B \wedge C) = f(B) \wedge f(C)$ ), and  $\Box$  is translated as *Prov* (so  $f(\Box B) = \text{Prov}(\ulcorner f(B) \urcorner)$ ). Thus Solovay’s arithmetical completeness theorem gives an alternative way to construct many nonprovable sentences. For example, we know that *PRL* does not prove  $\Box p \vee \Box \neg p$ , so by the theorem, there is an arithmetical sentence  $f(p)$  such that Peano arithmetic does not prove  $\text{Prov}(\ulcorner f(p) \urcorner) \vee$

$Prov(\neg f(p))$ ). In particular, if we suppose that Peano arithmetic does not prove any false sentences, this implies that neither  $f(p)$  nor  $\neg f(p)$  is provable in Peano arithmetic.

In recent years, logicians have investigated many other systems of arithmetic that are weaker than Peano arithmetic. They have given a partial answer to the question: “For which theories of arithmetic does Solovay’s arithmetical completeness theorem still hold?” It certainly holds for theories  $T$  that satisfy the following two conditions:

1.  $T$  proves induction for formulas in which all quantifiers are bounded (like the quantifier  $\forall x \leq y + z$ ) and  $T$  proves that for all  $x$ , its power  $2^x$  exists. In more technical terms:  $T$  extends  $I\Delta_0 + EXP$ .
2.  $T$  does not prove any false  $\Sigma_1$  sentences.

For such theories, it is also clear that  $PRL$  is sound if we read  $\Box$  as  $Prov_T$  (where  $Prov_T$  is a natural provability predicate with respect to a sufficiently simple axiomatization of  $T$ ). To sum up, we have the following theorem: If  $T$  satisfies 1 and 2, and  $A$  is a modal sentence, then

$$PRL \vdash A \Leftrightarrow \text{for all translations } f, T \vdash f(A).$$

This result shows a strength of provability logic: For many different theories,  $PRL$  captures exactly what those theories say about their own provability predicates. At the same time this is of course a weakness: For example, provability logic does not point to any differences between those theories that are finitely axiomatizable and those that are not.

In order to be able to speak in a modal language about such distinctions between theories, researchers have extended provability logic in many different ways, only a few of which are mentioned here. One way is to add a binary modality,  $\triangleright$ , where for a given theory  $T$ , the modal sentence  $A \triangleright B$  stands for “ $T + B$  is interpretable in  $T + A$ .” It appears that the interpretability logic of  $I\Delta_0 + superexp$  is different from the interpretability logic of Peano arithmetic.

Another way to extend the framework of  $PRL$  is to add propositional quantifiers, so that one can express principles like Goldfarb’s:

$$\forall p \forall q \exists r \Box((\Box p \vee \Box q) \leftrightarrow \Box r).$$

Finally, one can of course study predicate provability logic. V. A. Vardanyan proved that the set of always provable sentences of predicate provability logic is not even

recursively enumerable, so it has no reasonable axiomatization.

*See also* Gödel, Kurt; Gödel’s Incompleteness Theorems; Kripke, Saul; Logic, History of; Modal Logic; Peano, Giuseppe.

### Bibliography

- Artemov, S. N., and L. D. Beklemishev. “On Propositional Quantifiers in Provability Logic.” *Notre Dame Journal of Formal Logic* 34 (1993): 401–419.
- Boolos, G. *The Logic of Provability*. New York: Cambridge University Press, 1993.
- de Jongh, D. H. J., M. Jumelet, and F. Montagna. “On the Proof of Solovay’s Theorem.” *Studia Logica* 50 (1991): 51–70.
- Smoryński, C. *Self-Reference and Modal Logic*. New York: Springer-Verlag, 1985.
- Solovay, R. M. “Provability Interpretations of Modal Logic.” *Israel Journal of Mathematics* 25 (1976): 287–304.
- Visser, A. “Interpretability Logic.” In *Mathematical Logic (Proceedings, Chaika, Bulgaria, 1988)*, edited by P. P. Petkov. New York: Plenum Press, 1990.

*Rineke Verbrugge (1996)*

## PROVIDENCE

The idea of “providence” has three components—fore-sight, direction, and care. It is normally found in a theistic context. In its fullest sense it means that God foresees and governs (in a word, “provides for”) the world that is the object of his care (or love). Divine providence was affirmed by Plato in his *Laws* (887–888), where he condemns the view, later held by the Epicureans, that the gods take no interest in human affairs. The most important later thought upon the subject arose in Stoicism and Christianity.

### STOICS

The Stoics held a firm belief in the providence (*pronoia*) of God (or the gods). Thus, Epictetus uses an elementary form of the teleological argument to prove God’s supervision of the universe (*Discourse* 1.16). But two factors prevented the Stoics from taking a fully personal view of providence. First, they often conceived God abstractly (as a cosmic logos) and even physically (when they identified him with nature’s basic elements, air and fire). Second, and correlatively, they did not stress God’s care for persons individually, nor, as a consequence, did they allow that God accomplishes his purpose in and through the free response of human wills to his initiative. On the contrary, they equated providence with destiny or fate

(*heimarmene*). In the words of Cleanthes's Hymn to Zeus, translated by Seneca, *Ducunt volentem fata, nolentem trahunt* ("Fate leads the willing, drags the unwilling on," Epistles 107:11).

## CHRISTIANITY

Our primary evidence for Christianity is the teaching of Christ himself. Christ taught that God is a Father who cares for all his children individually. Therefore, they must not be anxious or distressed; rather, they must trust God absolutely (Matthew 6:25–33, 10:29–31). Furthermore, they must approach God freely in prayer in the confidence that he will answer their requests (Matthew 7:7–11). St. Paul made two basic assertions: first, that we know through Christ that God's sovereignty is one of love through which we are "more than conquerors" (Romans 8:35–39) and second, that God accomplishes his purpose by cooperating with our wills, not by demanding our submission to a fait accompli (Romans 8:14–16, Philippians 2:12–13). Hence, St. Paul, like Jesus, affirms the reality of, and the necessity for, petitionary prayer.

Attempts have been made to see providence in nature, history, and individual lives.

## NATURE

The theist maintains that God acts in nature both ordinarily, through those laws which science formulates, and extraordinarily, through miracles. Both modes of God's activity signify his wisdom and love to the believing mind. Furthermore, many theists, following Thomas Aquinas in his Fifth Way, believe that it is possible to base an argument for God's existence on the apparent traces of design in nature, but it must be admitted that the fact of evil constitutes prima-facie evidence against the existence of a Designer who is both omnipotent and good.

## HISTORY

To what extent can we interpret God's purpose in terms of a "pattern," or "patterns," discernible in historical events? Here one can only summarize a general tendency among modern theologians. Most of them would say that our ability to perceive a pattern or plan is restricted to the main events of the Bible as interpreted by the prophetic and apostolic writers. Perhaps we also have a right to see a *preparatio evangelica* in the achievements of Greece and Rome, but we cannot perceive an analogous plan in either the secular or ecclesiastical history of the postbiblical era. Thus, Josef Pieper writes, "Not that he who philosophizes could reach the point of being able to identify *in concreto*

the character of an event in terms of salvation and disaster. We are moving here within the realm of the *mysterious*—in the strictest sense. And even for the believer, the history of salvation 'within' history is not to be apprehended concretely" (*The End of Time*, London, 1954, p. 23).

## INDIVIDUAL LIVES

In regard to individual lives we must also distinguish between a general belief in providence and a detailed knowledge of its workings. St. Paul affirmed as a matter of faith that "we know that in everything God works for good with those who love him, who are called according to his purpose" (Romans 8:28). But in 1 Corinthians 13:12 he admits that all our knowledge of God is indirect, partial, and confused. Hence, any claim to see God's purpose in particular events is bound to be provisional and incomplete.

*See also* Christianity; Cleanthes; Epictetus; Epicureanism and the Epicurean School; Philosophy of Religion, History of; Plato; Seneca, Lucius Annaeus; Stoicism; Teleological Argument for the Existence of God; Thomas Aquinas, St.

## Bibliography

- Caussade, J. P. de. *L'abandon à la providence divine*. 1867. Translated by A. Thorold as *Self-Abandonment to Divine Providence*, 5th ed. London, 1955. A classic of spiritual theology.
- Flint, Thomas. *Divine Providence: The Molinist Account*. Ithaca, NY: Cornell University Press, 1998.
- Geach, Peter. *Providence and Evil*. Cambridge, U.K.: Cambridge University Press, 1977.
- Hazelton, Roger. *Providence*. London, 1958.
- Laird, John. *Mind and Deity*, 173–201. London: Allen and Unwin, 1941.
- Pollard, William C. *Chance and Providence*. New York: Scribners, 1958.
- Swinburne, Richard. *Providence and the Problem of Evil*. Oxford: Clarendon Press, 1998.

H. P. Owen (1967)

*Bibliography updated by Christian B. Miller (2005)*

## PSEUDO-DIONYSIUS

The writings of Pseudo-Dionysius, first cited at the beginning of the sixth century, have attracted interest partly because the writer has been wrongly identified with Dionysius the Areopagite, who was converted by St. Paul at Athens, and also with St. Denis, the patron saint of



France. Neither of these identifications, however, is possible.

While the thought of Pseudo-Dionysius was a continuation of the Christian Platonism of the early Church Fathers, it is directly influenced by the latest forms of Neoplatonism, as found in Proclus. No other early Christian writer was so clearly influenced by a particular philosopher. The influence of Pseudo-Dionysius on later theologians, philosophers, mystics, and poets was immense. John of Damascus and Thomas Aquinas were both strongly influenced by him. Peter Lombard, Robert Grosseteste, and Albert the Great also acknowledged their debt to him. The poetry of Dante Alighieri and John Milton reflects his heavenly hierarchy.

Four of his treatises—"The Celestial Hierarchy," "The Ecclesiastical Hierarchy," "The Divine Names," and "The Mystical Theology"—and ten of his letters are extant. The problem of the one and the many in the treatises is the problem of the relation of God to the universe, both visible and invisible. The basic propositions of Proclus were that every plurality participates in unity, is both one and not one, and is other than the one itself. The order of the universe is an order that depends on the ultimate unity. It is arranged in different orders of being that descend from and ascend to the first principle. This hierarchical view of the universe goes back to Plato and Aristotle and is found in Philo and the Gnostics, as well as in later Platonism. Proclus and Pseudo-Dionysius represent the final stage of the idea in the ancient world, and Pseudo-Dionysius is the chief transmitter of the idea to later times.

The four treatises exhibit the sequence of Dionysius's thought. Those on hierarchies show the descent and return of the divine goodness, "The Divine Names" shows the nature of God, and "The Mystical Theology" shows the way by which the knowledge of God may be found.

## THE HIERARCHIES

"The hierarchy is a holy order, a knowledge and an activity which assimilates to the divine nature as far as possible and which through the light granted from God is raised in due proportion to the imitation of God" ("The Celestial Hierarchy" III, 1). The celestial hierarchy contemplates the divine perfection and shares in it, reflecting its light down through its several ranks: Seraphim, Cherubim, Thrones, Dominions, Powers, Authorities, Principalities, Archangels, and Angels. The members of the highest hierarchy are nearest to God and share most fully his vision and his likeness. The other members of the hierarchy become more symbolic and corporeal as they

descend. Each member of the hierarchy comes directly from God, in contrast with the emanations of Proclus, which produce one another. The Christian doctrine of creation makes the unity of the hierarchy that of spiritual communion rather than that of progressive generation. On Earth the ecclesiastical hierarchy continues the celestial hierarchy in visible form, with Jesus at the top of this hierarchy as God is at the summit of the celestial hierarchy. The members of the hierarchy in descending triads are chrism, communion, and baptism; bishops, priests, and deacons; monks, laity, and catechumens.

## "THE DIVINE NAMES"

The third treatise discusses the names given to God. These names cannot describe God but must be understood in a special sense, since God is above all reason, speech, being, and name. He is above being yet the cause of being, and may be said to be only in a higher sense. His names are not derived from himself but from the manifestation of his providence. He is both nameless and many-named. He is in the world, around the world, above the world, and above the heavens. He is sun, star, fire, water, wind, dew, cloud, stone, and rock—and none of them. Knowledge of God comes through prayer, which draws men to him so that they may know his goodness. How can such a God be the sovereign creator of a world in which evil exists? Only because evil is not real but simply the absence of good. "Evil is then a deprivation, defect, weakness, disproportion, error, and the absence of purpose, beauty, life, understanding, reason and perfection." When night falls, there is nothing positive in its darkness but simply the absence of light. Evil is simply the absence of goodness.

## "THE MYSTICAL THEOLOGY"

"The Mystical Theology" describes the way to the knowledge of God by the Neoplatonic method of abstracting visible and invisible qualities until one comes to the knowledge of God by negation or removal. This knowledge of God is mystical and ineffable rather than philosophical and theological, and it involves complete cessation of thought and speech. One penetrates the darkness that is above intelligible things, and in absolute silence one is united to the ineffable. God is absolutely unknowable, and the ecstasy that unites with him is both total ignorance and a knowledge beyond reason.

The distinctive quality of Pseudo-Dionysius is found in the extreme statement of two things—the unity of the world and the unity of God. The unity and order that the divine goodness imposes on the universe is described

most concretely and explicitly. The unity of God is described in negative terms that isolate it completely from all else. The extreme statement of these two opposite things enabled Pseudo-Dionysius to influence succeeding thinkers in their account of an ordered world and of a transcendent God. The opposition is inherent in all Platonism if not in all philosophy. Its explicit exposition is therefore of value.

**See also** Albert the Great; Aristotle; Dante Alighieri; Grosseteste, Robert; John of Damascus; Milton, John; Neoplatonism; Patristic Philosophy; Peter Lombard; Plato; Platonism and the Platonic Tradition; Proclus; Thomas Aquinas, St.

### Bibliography

#### WORKS BY PSEUDO-DIONYSIUS

- Migne, J. P., ed. *Patrologia Graeca*. Vol. III. Paris, 1857.  
 Suchla, Beate Regina Suchla, et al. *Corpus Dionysiacum*. 2 vols. Patristische Texte und Studien 33, 36, Berlin–New York: Walter de Gruyter, 1990–1991. Critical edition  
 Luibheid, Colm, tr. *The Complete Works*. Classics of Western Spirituality. Mahwah NY: Paulist Press, 1987. English translation.

#### WORKS ON PSEUDO-DIONYSIUS

- de Andia, Ysabel, ed. *Denys l'Aréopagite et sa postérité en orient et en occident*. Collection des ...tudes Augustiniennes, Série Antiquité 151. Paris: Institut d'Études Augustiniennes, 1997.  
 de Andia, Ysabel. *Henosis: L'union à Dieu chez Denys l'Aréopagite*. Leiden: E.J. Brill, 1996.  
 Louth, Andrew. *Denys the Areopagite*. London: Geoffrey Chapman, 1989.  
 Roques, René. *L'univers dionysien*. Paris: Aubier, 1954.  
 Rorem, Paul. *Pseudo-Dionysius. A Commentary on the Texts and an Introduction to Their Influence*. New York: Oxford University Press, 1993.

E. F. Osborn (1967)

Bibliography updated by Andrew Louth (2005)

## PSEUDO-GROSSETESTE

Pseudo-Grosseteste was the anonymous author of a *Summa Philosophiae*, written between 1265 and 1275. Because of the reference in the *Summa* to Simon de Montfort's death (1265), it could not have been written by Robert Grosseteste, who died in 1253. Bartholomew of Bologna, Robert Kilwardby, and a disciple of Roger Bacon have all been suggested as the author, but there is no consensus. It does seem probable, however, that he was English and was either a Franciscan or a secular.

The *Summa*, which begins with a history of philosophy similar to that found in Bacon's *Opus Maius*, is a work of considerable subtlety and sophistication, an advanced product of the so-called Augustinian school. It holds that there is a universal wisdom in which both ancients and moderns share, perfected however by Christian revelation. Those concerned with wisdom are theosophists, to whom truth is directly revealed; theologians, who systematize and make more clear what has been revealed to the theosophists; and philosophers. The first two groups are concerned with the infallibly true, and their proper study is of matters relevant to human salvation. Philosophy, on the other hand, while it may often be in error, is completely unrestricted in its scope and may undertake to explain the natures and causes of all things whatsoever.

The *Summa* then treats the whole range of metaphysical questions in separate treatises, beginning with truth and the necessary existence of an uncreated being and ending with psychology, light, the four elements, meteors, and minerals. Its characteristic metaphysical positions are derived largely from the author's explicit hylomorphism. Every created thing is composed of matter and form. Prime matter, the mark of contingency, is not corporeal but is unextended and has three inseparable properties: It is in potency to every form; it has a desire for form; and it is privation of form. Insofar as it is privation of form it is the cause of instability; but its desire for form is a tendency toward stability. It first receives universal form, that is, substance. Substance, or substantial form, is either corporeal or incorporeal and individuates matter. It receives further perfections from other forms, so that there is a plurality of forms in any given body. This leads the author to reject the distinction (except as one of reason) between essence and existence. It also leads him to insist that the Intelligences are compounded of matter and form and differ both according to species and individuality. The human soul, like the Intelligences, is an incorporeal intelligent substance, but unlike them is capable of being joined to a body as well as of existing separately; it too is composed of matter and form. In these points, as in many others throughout the *Summa*, the author seems to be correcting what he considers the errors of Thomas Aquinas.

**See also** Augustinianism; Bacon, Roger; Essence and Existence; Grosseteste, Robert; Kilwardby, Robert; Revelation; Thomas Aquinas, St.

### Bibliography

The text of the *Summa* is printed in L. Baur, *Die philosophischen Werke des Robert Grosseteste, Bischofs von Lincoln* (Münster in Westfalen: Aschendorff, 1912), pp. 275–643. Parts are translated in R. McKeon, *Selections from Medieval Philosophers* (New York: Scribners, 1929), Vol. I, pp. 290–314. An excellent analysis appears in C. K. McKeon, *The Summa Philosophiae of the Pseudo-Grosseteste* (New York, 1948). A good summary is in E. Gilson, *History of Christian Philosophy in the Middle Ages* (New York: Random House, 1955), pp. 265–274.

Richard C. Dales (1967)

## PSYCHĒ

“Psychē” in Homer first means life and later means a departed life or ghost. The first identification with soul in the sense of the conscious self is found perhaps in Ionia, and the earliest full identification with the rational as well as with the emotional side of personality has been attributed to Socrates. In all this there was no opposition between soul and body. The doctrine that the soul is a prisoner in the body that Plato took from Orphic doctrine had reached Greece, perhaps from Scythia, before the time of Pythagoras, probably in association with a doctrine of transmigration. Plato, in the *Phaedo*, while recognizing that most people do not believe in survival after death (80D), propounded a view that combines the Socratic and Orphic attitudes. In the tripartite soul of the *Republic*, however, it is the rational part alone that is immortal; this was also Aristotle’s view.

The majority of the pre-Socratics regarded the universe as a quasi-living organism, and this view also found expression in Plato’s doctrine in the *Timaeus* of a world soul as a source of orderly motion in the universe. Aristotle presented a developed human and animal psychology in his analysis of the soul in the *De Anima* and elsewhere. Whereas Plato regarded the soul as a substance separate from the body, Aristotle’s final view treated it as the form of a living body. For the Stoics the soul is an aspect of the all-pervading cosmic logos, while for the Epicureans it is a combination of especially smooth atoms. Within Christian theology Augustinians follow an essentially Platonist view, while Thomists prefer Aristotle’s approach.

**See also** Aristotle; Augustinianism; Epicureanism and the Epicurean School; Homer; Orphism; Plato; Platonism and the Platonic Tradition; Pre-Socratic Philosophy; Pythagoras and Pythagoreanism; Socrates; Stoicism; Thomism.

### Bibliography

Burnet, J. “The Socratic Doctrine of the Soul.” *Proceedings of the British Academy* 7 (1915–1916): 235–259.  
Rohde, E. *Psyche*. Translated by W. B. Hillis. London: Routledge and K. Paul, 1925.

G. B. Kerferd (1967)

## PSYCHOANALYSIS

The term *psychoanalysis* pertains to the theory, therapy, and method of inquiry created by Sigmund Freud (1856–1939). The origin of psychoanalysis is often traced to Freud’s *Interpretation of Dreams* (1900), but some of its key elements can be found in his earlier writings, especially his *Studies on Hysteria* (1895), cowritten with Josef Breuer.

### FREUDIAN THEORY

“Freudian theory” is not a single theory but a set of smaller ones, at least some of which are familiar to most philosophers and educated people. One of the most significant and best known of these is Freud’s theory of dreams.

Freud theorized that all dreams are fulfillments of repressed infantile wishes. During sleep, these repressed wishes can enter into the dreamer’s consciousness, but only in a disguised form, after the dream censor has altered their appearance. Freud calls what survives the dream censorship the “manifest content”; what exists prior to the censorship is the dream’s “latent content.” By having a patient free associate to a dream’s manifest content, Freud hoped to determine the dream’s latent content and ultimately to glean information about a patient’s unconscious conflicts.

In *The Interpretation of Dreams*, Freud distinguishes three areas of the human mind: consciousness, the preconscious, and the unconscious. Consciousness contains all that we are immediately aware of; the preconscious contains mental contents that we can easily become aware of; and the unconscious contains mental contents that cannot be brought to consciousness except through the use of psychoanalytic therapy.

The reason that unconscious ideas cannot readily be brought to consciousness, according to Freud, is that they are repressed. Repression and the unconscious are closely linked in his early writings: “Thus we obtain our concept of the unconscious from the theory of repression. The repressed is the prototype of the unconscious for us” (Freud 1923, p. 15). On his early theory of the dynamic

unconscious, what is repressed is unconscious and what is unconscious is repressed.

On September 26, 1922, however, Freud read a short paper at the Seventh International Psycho-Analytical Congress, “Some Remarks on the Unconscious,” in which he indicated dissatisfaction with his theory. In an abstract of the paper (the abstract may have been written by Freud himself), it is noted that the speaker (i.e., Freud) had retold the history of the development of the theory of the unconscious and had pointed out that it had been deemed necessary to equate the dynamic unconscious with the repressed. “It has turned out, however, that it is not practicable to regard the repressed as coinciding with the unconscious and the ego with the preconscious and conscious. The speaker discussed the two facts which show that in the ego too there is an unconscious, which behaves dynamically like the repressed unconscious...” (author unknown, 1923, p. 367). The two facts are resistance proceeding from the ego during analysis and an unconscious sense of guilt.

This short paper and its abstract anticipated the publication of his *The Ego and the Id* (1923), in which Freud makes another important modification of his earlier views. Here he introduces the expression “*das Es*” (“the it”), which he explicitly borrows from Georg Groddeck; it has been translated by Freud’s English translators as “the id.” On his new theory, the structural theory, the unconscious is not equated with the repressed. All that is repressed is unconscious, but some of what is unconscious is not repressed. Some of what is in the id is repressed, but some of it is not. In addition, Freud now divides the mind into the id, ego, and superego.

The ego is held to be partly conscious and partly unconscious. It negotiates the demands of the outside world and those of the id and the superego. The id is largely unknowable, according to Freud, but we can know that it exists and know some of its properties. The id is entirely unconscious; it seeks satisfaction only of its instinctual needs, and it is the source of much psychic conflict. The superego develops out of the ego and maintains a system of ideals, values, and prohibitions.

At first, Freud tended to equate repression and defense, but in later works he classifies repression as but one type of defense. Other defense mechanisms include projection, reaction formation, sublimation, isolation, and regression. Despite his work on these other types of defenses, Freud still held that repression was the most important type of defense. In fact, he saw repression as the “cornerstone” of the whole structure of psychoanalysis.

Freud appealed to repression as the important causal determinant of parapraxes, which include memory mistakes, slips of the tongue, and neuroses, although not all neuroses. What Freud called “actual neuroses,” including anxiety neuroses and neurasthenia, are caused by events in later life and are not explainable by Freudian theory. What Freudian theory does purport to explain are the “psychoneuroses,” such as obsessional neurosis, hysteria, and depression. The psychoneuroses are said to arise from the repression of erotic wishes; their symptoms are “compromise formations”—they represent a solution to unconscious conflicts among the id, ego, and superego.

Another significant Freudian theory concerns sexual stages of development. Each of us, it is theorized, goes through four such stages. In the first year, the infant passes through the oral stage, during which its mouth is its primary source of pleasure. The focus then changes in the anal stage, where, during the next three years or so, the infant’s interest shifts to its anus. From three to five years, the child passes through the phallic period, and its genitals are of major interest. There is then a latency period lasting until puberty, when an interest in sex reemerges.

How a child reacts to events during the various stages of sexual development can help determine its adult personality. Both Freud and his followers theorized that certain personality clusters, such as obstinacy, parsimony, and orderliness, were causally linked to specific events in one or other of the infantile stages of development.

In addition to the four stages of sexual development, Freud postulated another stage, the oedipal phase, lasting roughly from age three to five years. During this period, the male child unconsciously desires to possess his mother sexually, but because of perceived threats from his father, the child develops what Freud terms the “oedipal complex.” The boy begins to fear that his father will cut off his penis and develops castration anxiety. Freud’s first published discussion of castration anxiety occurs in his discussion of the case of Little Hans (1909), whose mother told him that if he continued to touch his penis, she would ask the doctor to cut it off.

Some Freudians postulate in little girls a complex analogous to the oedipal complex, the “electra complex.” It was Carl Jung, however, not Freud, who introduced this concept. Freud himself doubted that the concept was useful and even that the phenomenon occurred (Freud 1931, p. 229).

## PSYCHOANALYTIC THERAPY AND METHOD OF INQUIRY

Standard psychoanalysis, or “analysis,” has certain features that distinguish it from other types of psychotherapy. The analysand, the patient, reclines on a couch, while the analyst remains out of sight. The therapy is scheduled for four or five times per week, and, in contrast to short-term psychoanalytically oriented psychotherapy, it typically lasts three years or more.

Some analysts distinguish three phases of an analysis: the beginning, middle, and end. In the beginning phase, the analyst has a preliminary consultation with the patient, sets the fee schedule, determines (in consultation with the patient) the days and hours of analysis, and decides whether the patient is a suitable candidate for analysis. The initial phase of the analysis can last from a week or two to several months.

In the middle phase, which can last years, the heart of the analysis takes place. The patient is instructed to report childhood memories and dreams and to free-associate about their contents, saying whatever comes to mind without pause or hesitation. The main methods of inquiry used by the analyst are free association and interpretation. The analyst uses the observed data from the patient’s free associations in conjunction with other observations from the therapy sessions to form preliminary interpretations, or hypotheses, about the cause of the patient’s problems. These hypotheses are modified as the analysis progresses, with the analyst taking into account some very important factors that tend to emerge later in the analysis.

One of these factors is resistance. Freud explained the resistance that eventually emerges as the attempt to defend against remembering what has been repressed. Resistance can take different forms, including certain verbalizations, expressions of recalcitrant attitudes, and the unwillingness to free-associate. In his *Inhibitions, Symptoms, and Anxieties* (1926 [1925], p. 159), Freud refers to the patient’s resistance as “the resistance of the unconscious.”

The solution to the resistance problem is to let the patient “work through” the resistances:

One must allow the patient time to become more conversant with this resistance with which he has now become acquainted, to *work through* it, to overcome it, by continuing, in defiance of it, the analytic work according to the fundamental rule of analysis. Only when the resistance is at its height can the analyst, working in common

with his patient, discover the repressed instinctual impulses which are feeding the resistance; and it is this kind of experience which convinces the patient of the existence and power of such impulses. The doctor has nothing else to do than to wait and let things take their course, a course which cannot be avoided nor always hastened” (1914, p.155).

Another significant factor, one of “undreamt-of importance” (Freud 1940, p. 174), is transference. In the course of the analysis, the patient comes to see the analyst as the reincarnation of some important figure in his or her past and “transfers” to the analyst the negative or positive feelings formerly directed to the figure from the past. An important part of the analysis consists of the analyst’s attempt to analyze the overt manifestations of the patient’s transference in order to reach a final interpretation of the patient’s problems.

In the third and last phase of the analysis, the final interpretation is revealed to the patient: The repressed is made conscious. Yet no mere telling of the interpretation is likely to have any lasting therapeutic effect unless the ego has been strengthened enough to enable the patient’s acceptance of the interpretation.

In *Analysis Terminable and Interminable* (1937), Freud gives two criteria for terminating the analysis: first, symptom relief, with the patient overcoming his anxieties and inhibitions; and, second, the analyst’s judging that so much material has been made conscious and so much resistance conquered that there is no need to fear a repetition of the pathological processes that caused the patient’s problems. These criteria are relevant to deciding in *one sense*, Freud says, if there is to be “the end of an analysis,” but in another sense, more is required. In asking whether the analysis is at an end in this second sense, we are asking whether the analyst has had such a far-reaching influence on the patient that no further change could be expected to take place in him if his analysis were to continue. “It is as though,” Freud writes, “it were possible by means of analysis to attain to a level of absolute psychical normality,” as though the analyst had succeeded in resolving every one of the patient’s repressions (1937, pp. 219–220).

The above material contains the main outlines of Freud’s most important theories, his method of inquiry, and his therapy, but not all of his theories are covered, and important details are necessarily omitted. For brief discussions of additional psychoanalytic concepts, see B. Moore and B. Fine (1990); for more detailed discussions

of nearly all of Freud's theories, the history of the psychoanalytic movement and its development in countries around the world, and the contributions of other major figures to the development of psychoanalysis, see E. Erwin (2002).

### FISSURES IN THE MOVEMENT

In the early years of the psychoanalytic movement, two serious schisms occurred: Alfred Adler (1870–1937) broke with Freud in 1911 and, at approximately the same time, Carl Jung (1875–1961) began fighting with Freud and in 1914 resigned from the International Psychoanalytical Association. These figures disagreed with Freud about several matters, but especially about the theoretical importance placed by Freud on infantile sexuality.

After breaking with Freud, Adler went on to develop his own general psychology. One of his key ideas is that the psychologically disturbed individual suffers from extreme feelings of inferiority. One of the main goals of Adlerian therapy is to eliminate this feeling of inferiority and to put in its place a feeling of community and connectedness with others. Carl Jung also developed his own type of psychotherapy and along with it a rich and complex theoretical framework that included the postulation of the collective unconscious, his theory of archetypes, and his distinction between “extroverts” and “introverts.”

One could view the theorizing of Adler and Jung as taking psychoanalysis in new directions, but their theories are so radically different from Freud's that it is doubtful that either's theory or therapy is a form of psychoanalysis at all. When Adler left—or rather was pushed out of—the Vienna Psycho-Analytical Society, he started his own group, “The Society for Free Psycho-Analysis,” but he quickly changed the name of his theory to “Individual Psychology,” a step for which, Freud said, “we are all thankful.” (“There is room enough on God's earth, and anyone who can has a perfect right to potter about on it without being prevented; but it is not a desirable thing for people who have ceased to understand one another and have grown incompatible with one another to remain under the same roof.” [Freud 1914, p. 52]). Jung, like Adler, also did not characterize his theory or therapy as a form of psychoanalysis; he preferred the name “analytical psychology.”

Long after the departure of Adler and Jung, other cracks developed in the psychoanalytic movement, but these were much smaller. One of the first of these resulted from the work of Melanie Klein, a Budapest psychoanalyst who in 1926 moved to London, where she continued her work analyzing children. Klein saw herself as contin-

uing Freud's work, although she did depart from his theories in certain respects, such as postulating the occurrence of oedipal conflicts in little girls and at an earlier time than specified by Freud's theory. Klein claimed to have made a series of important discoveries about infants, including their having a terrifying mental life, populated by beasts and monsters, and having cannibalistic urges causally linked to earlier contact with the mother's breast. Anna Freud, also working in London at the same time, strongly disagreed with some of Klein's theorizing and managed to win the support of the Vienna Psycho-Analytical Society in condemning Klein's views. The result was a bitter dispute between the “Kleinians” and London psychoanalysts who sided with Anna Freud.

A second division occurred because of the development of ego psychology, the groundwork for which was laid first by Freud's *The Ego and the Id* (1923) and developed further by Anna Freud in her work *The Ego and the Mechanisms of Defense* (1946 [1936]). Ego psychology began to flourish within the psychoanalytic tradition with the publication of Heinz Hartmann's *Ego Psychology and the Problem of Adaptation* (1958 [1939]). Hartmann and his colleagues did not see themselves as breaking with the Freudian tradition in any serious way, but they placed far more emphasis than did Freud on the role of the ego, while greatly reducing the theoretical significance of the id and superego.

Two further theoretical sharp turns occurred in the second half of the twentieth century with the development of object-relations theory and self psychology, now two of the most dominant forms of psychoanalysis.

Object-relations theory developed out of the work of British psychoanalysts, among them Melanie Klein, W. D. Fairbairn, and D. W. Winnicott. This theory is also associated with the work of psychoanalysts living in the United States, such as Otto Kernberg. According to traditional Freudian theory, there exists in each individual biological, instinctual urges, the mental representation of which are referred to as “drives.” There are two sorts of drives: the sexual drive and the drive for self-preservation. Object-relations theorists reject Freud's biologically oriented drive theory and argue that the infant is motivated not by instinctual urges but by the need to relate to another person, such as the mother. Freud, like the object-relations theorists, also used the term *object* in his discussion of infants, but he was referring not to people or things external to the infant but to the child's mental representation of them.

In contrast, object-relations theorists tend to refer to things or persons in close proximity to the infant as

“objects,” but, somewhat confusingly, the theory also talks of “internalized objects,” which clearly are not objects in the external world. One leading theorist, W. D. Fairbairn, in his *Psychoanalytic Studies of the Personality* (1952, p. 137) distinguishes between “objects” and “internalized objects” in terms of a contrast between normal and pathological psychology. In the object-relations theory, psychology becomes “the study of the relationships of the individual to his objects, whilst, in similar terms, psychopathology may be said to resolve more specifically into a study of the relationships of the ego to its internalized objects” (Fairbairn, 1952, p. 137).

In self psychology, the key theoretical concept, that of a “self object,” also has a double use; it is sometimes applied to persons and at other times to their mental representations. Self psychology was developed by Heinz Kohut and his colleagues. Kohut became known for his theory of the narcissistic personality disorder, said to have a different etiology from the “transference neuroses” talked about by Freudian theory. This disorder, Kohut claimed, can be recognized partly by observing its distinct symptoms and partly by analyzing the different types of transference that develop in the course of the analysis: a mirroring, idealizing, and twinship transference. Each of these transferences reflects the failure of a parent to respond adequately to a different type of need of the infant, such as the child’s need to confirm its own sense of greatness (the need for a “mirroring” response) or the need to experience others who resemble it (the need for a “twin” response). The result of these failures to respond is the narcissistic pathology, the subsequent failure of the narcissistic person to develop an intact self.

## THE CURRENT STATUS OF PSYCHOANALYSIS

Freud’s theorizing has had an enormous influence on psychiatry, clinical psychology, art, cinema, literature, religion, anthropology, history, biography, sociology, and philosophy. The remnants of his theorizing survives through the work of individual psychoanalysts and the work of the psychoanalytic institutes and associations that exist in the United States, Great Britain, Brazil, Sweden, Finland, Mexico, South Africa, France, Austria, and in many other countries.

The work of the breakaway theorists, Alfred Adler and Carl Jung, has been considerably less popular than Freudian theory, but their theories have nonetheless been influential and are still accepted by many. Many Adlerians belong either to the International Association for Individual Psychology or to the North American Society of

Adlerian Psychology. There are also Alfred Adler institutes and schools in Chicago, San Francisco, Washington, New York, and other cities.

Many adherents of the theories of Jung belong to the International Association for Analytical Psychology. C. G. Jung institutes and societies are located in this country in New York, Seattle, Portland, Boston, Los Angeles, and other large cities, and in Canada, Australia, Great Britain, and other countries.

The continued influence of various psychoanalytic theories is important, but there is also the question of truth: How much of psychoanalytic theorizing is at least approximately true? Some of the things that Adler and Jung said were rather commonsensical and not controversial or original. If we subtract these propositions, how many of their distinctive and original claims have been shown to be true? Not very many. There are few, if any, formal empirical studies of their theories. The verdict must be that their theories remain little more than interesting but unproven conjectures.

The work of the ego psychologists, the object-relations theorists, and the self psychologists has been the subject of more empirical inquiry, but there is nothing that can be said to constitute a firm body of supporting evidence for any one of these modifications of Freudian theory. This fact has led one prominent psychoanalyst to point out that the developments in ego psychology were not prompted by new data in the psychoanalytic situation but by the recognition of obvious deficiencies in Freudian theory, and that none of these three theories has remedied the epistemological and methodological difficulties associated with Freudian theory (Eagle 1993).

The sheer quantity of the empirical evidence for Freudian theory and therapy is far greater than that of its newer psychoanalytic rivals. It includes not only Freud’s case studies but also the published case studies of many of his followers, data from anthropology and the “psychopathology of every day life,” and more than 1,500 experimental studies. There are also Freud’s arguments to consider: They are designed to show that even without the benefit of controlled studies, his theories receive powerful support from the data obtained from psychoanalytic case studies.

In evaluating the Freudian evidence, one issue concerns its subject. There is a watered-down, commonsensical version of Freud’s theories and there are the original, distinctively Freudian versions articulated and modified over the years principally by Freud himself. On the watered-down version, the unconscious exists if a person

has mental states that exist below the threshold of consciousness, whether or not these states can be brought to consciousness without the aid of psychoanalysis. Repression is said to occur whenever one tries to keep something painful out of consciousness, which obviously happens when one tries to forget a sad love affair or a hurtful insult. There are “Freudian slips,” it is said, if people make linguistic mistakes with sexual innuendoes, regardless of what causes the errors. Defense mechanisms such as “projection,” “reaction formation,” and “displacement” are said to be operative so long as certain types of defensive behavior are displayed, such as attributing to others one’s own faults or doing just the opposite of what one would like to do, no matter what causal mechanism explains the behaviors.

The evidence for some of the best-known hypotheses of the popularized, watered-down version of Freudian theory is quite strong but not new: the evidence for some sort of unconscious mind, intentional forgetting, slips of the tongue, and defensive behaviors was known to psychologists and philosophers of the nineteenth century, before Freud invented psychoanalysis. Recent historical research has shown that even many of Freud’s seemingly distinctive ideas were anticipated, not merely in some vague way but in detail, by the philosophers Arthur Schopenhauer (1788–1860), Friedrich Nietzsche (1844–1900), Eduard von Hartmann (1842–1906), and J. F. Herbart (1776–1844) (Zentner 2002).

If we limit the discussion to what is distinctively Freudian, scholars still disagree about what the evidence shows. Some still claim that the evidence gleaned from clinical case studies strongly supports some parts of Freudian theory, although that view is losing adherents even among Freudians, partly or largely due to the trenchant and systematic criticisms of the Freudian clinical evidence by the philosopher Adolf Grünbaum (2002). If, as Grünbaum argues, the clinical evidence has little probative value with respect to Freudian theory, that leaves mainly the Freudian experimental evidence, said by some to firmly support some central parts of the Freudian corpus (Kline 1981, Fisher and Greenberg 2002). Another review of the very same experimental evidence concludes that it provides almost no support for any distinctively Freudian hypothesis (Erwin 1996).

As regards Freud’s therapeutic claims, there are uncontrolled case studies and correlational studies of long term orthodox psychoanalytic therapy, but there has never been a randomized clinical trial studying its effects. Two retrospective studies of long-term psychoanalysis have been published in recent years; some analysts argue

that despite their lack of controls, they provide support for the effectiveness of psychoanalysis because of the employment of novel statistical techniques or the presence of other features that obviate the need for experimental controls. These studies and the claims on their behalf are criticized in Erwin (2002).

*See also* Freud, Sigmund.

## Bibliography

### WORKS BY FREUD

- All references to Freud’s works are to *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. 24 vols. Translated by James Strachey. London: Hogarth Press, 1953–1974.
- Breuer, J., and S. Freud. *Studies on Hysteria* (1895). *S. E.* 2, 19–305.
- The Interpretation of Dreams* (1909). *S. E.* 4 and 5: 1–627.
- “Analysis of a Phobia in a Five-Year-Old Boy” (1909). *S. E.* 10: 5–149.
- “On the History of the Psycho-Analytic Movement” (1914). *S. E.* 14: 3–86.
- The Ego and the Id* (1923). *S. E.* 19: 12–63.
- Inhibitions, Symptoms and Anxiety* (1926). *S. E.* 20: 87–172.
- “Female Sexuality” (1931). *S. E.* 21: 225–243.
- Analysis Terminable and Interminable* (1937). *S. E.* 23: 216–253.
- An Outline of Psychoanalysis* (1940). *S. E.* 23: 174.

### WORKS ABOUT FREUD

- Freud, A. *The Ego and the Mechanisms of Defense*. New York: International Universities Press, 1946 [1936].
- Author Unknown. Abstract. *Int. J. Psycho-Anal.*, 1923, 4: 367.
- Eagle, M. “The Dynamics of Theory Change in Psychoanalysis.” *Philosophical Problems of the Internal and External Worlds: Essays on the Philosophy of Adolf Grünbaum*, edited by J. Earman, A. Janis, G. Massey, and N. Rescher. Pittsburgh: Pittsburgh University Press, 1993.
- Erwin, Edward. *A Final Accounting: Philosophical and Empirical Issues in Freudian Psychology*. Cambridge, MA: MIT Press, 1996.
- Erwin, Edward. *The Freud Encyclopedia: Theory, Therapy, and Culture*. New York: Routledge, 2002.
- Fairbairn, W. D. *Psychoanalytic Studies of the Personality*. London: Tavistock Publications and Routledge & Keegan Paul, 2002.
- Fisher, S., and R. Greenberg. “Scientific Tests of Freud’s Theories and Therapy.” In *The Freud Encyclopedia: Theory, Therapy, and Culture*, edited by Edward Erwin. New York: Routledge, 2002.
- Grünbaum, A. “Critique of psychoanalysis.” In *The Freud Encyclopedia: Theory, Therapy, and Culture*, edited by Edward Erwin. New York: Routledge, 2002.
- Grünbaum, A. *The Foundation of Psychoanalysis: A Philosophical Critique*. Berkeley: University of California Press, 2002.
- Hartmann, H. *Ego Psychology and the Problem of Adaptation*. New York: International Universities Press, 1958 [1939].



Kline, P. *Fact and Fantasy in Freudian Theory*. London: Methuen, 1981.

Moore, B., and B. Fine, eds. *Psychoanalysis: The Major Concepts*. New Haven: Yale University Press, 1995.

*Edward Erwin (2005)*

## PSYCHOANALYSIS, EXISTENTIAL

See *Existential Psychoanalysis*

## PSYCHOANALYTIC THEORIES, LOGICAL STATUS OF

Since psychoanalysis fails to conform to currently accepted methodological models, its prominence on the contemporary scene constitutes a challenge to the methodologist. He must either revise his canons or show the psychoanalyst the error of his ways. Both tacks have been tried, but thus far the second has predominated. This entry will be confined to methodological problems raised by psychoanalytic theory, though as we shall see, such problems cannot be pursued very far without running into questions concerning the clinical interpretation of particular cases.

### CONTENT OF PSYCHOANALYTIC THEORY

Within psychoanalytic theory there are diverse strands, and the relations between them are by no means obvious. For one thing, there are theoretical ideas at different levels. Fairly close to actual clinical practice are found the concepts of repression, regression, projection, reaction formation, and transference. At a higher level there is a theoretical model of the mind in terms of psychic energy, which gets attached to various ideals, the transformations of which are governed by quasi-mechanical principles. This is, in fact, designed to be a perfectly general model of the mind, in terms of which, in the last analysis, all psychological processes and states may be conceived. At this level we have also the division of the psyche into the three systems—id, ego, and superego—together with an account of their properties and interrelations.

In addition to the distinction between levels, we have the distinction between developmental and dynamic theories. In the first group is the theory of psychosexual

stages—oral, anal, genital—according to which there is a biologically determined order, beginning from infancy, in which first one, then another, area of the body is maximally sensitive to pleasurable stimulation and according to which certain personality traits predominate as one or another stage is prolonged or transcended only with difficulty. For example, passivity and lack of initiative are associated with the oral stage, during which sensuous pleasure comes mostly from taking things into the mouth.

By contrast, the dynamic theories have to do with processes that take place, or can take place, over a short span of time or at least within the same stage of a person's life. Under this heading we have, for example, the theory of defense mechanisms, according to which the person will defend himself against dangerous impulses by various devices—going to the other extreme (reaction formation), attributing the impulses to someone else (projection), and so on. One of the reasons that the distinction between developmental and dynamic theories is important is that many of the philosophical difficulties raised about psychoanalytic theory center on the notion of unconscious psychic processes, and such processes are more central in dynamic than in developmental theories.

In order to have something fairly definite to work with, let us take the following to be an oversimplified formulation of the psychoanalytic theory of psychic conflict, which is basic to all the dynamic theories.

(1) When it is very painful for a person to be aware of the fact that he has a certain desire, he represses it (prevents it from becoming conscious). The pain may stem from a severe conflict between the desire and the person's standards for himself, from fear of the consequences of attempts to satisfy the desire, or from both.

(2) Repressed psychic material exhibits primitive, infantile features. These include the lack of sharp distinctions, which is in turn conducive to the formation of strong associations between a certain desire and many other, often irrelevant, things and a tolerance for lack of realism and for incompatibility of one's desires and thoughts.

(3) A repressed desire (which continues to exist as a desire) can be partially satisfied by happenings, in actual occurrence or in fantasy, which are associated with the object of the desire.

(4) When the substitute satisfactions themselves arouse too much anxiety, the person seeks to ward them off, often in equally derivative ways.

This basic theory is then applied to the explanation of dreams, slips of the tongue, and neurotic symptoms by studying the ways in which such phenomena constitute substitute satisfactions of repressed desires and/or defenses against such satisfactions.

An illustration of these ideas is presented by Sigmund Freud in Lecture 17 of his *General Introduction to Psychoanalysis*. A girl has, for obvious reasons, repressed a strong desire for sexual intercourse with her father. In the unconscious, various things happen to this desire and the ideas involved in it. The dread of carrying out the act generalizes to a dread of sexual activity of any sort. An association is formed between sexual intercourse and breaking a vase. The bolster at the back of the bed is pictured as the girl's father and the back of the bed as her mother. The pressure of this repressed material becomes so great that the girl develops a compulsion to go through an elaborate ritual before going to sleep at night. She arranges the vases in her room so that breakage is impossible, thus symbolically guarding against sexual intercourse, and she takes care lest the bolster touch the back of her bed, thus achieving a substitute satisfaction for her desire to keep her father and mother apart.

### METHODOLOGICAL PROBLEMS

Some of the philosophical objections to psychoanalytic theory can easily be shown to have little or no force. For instance, some philosophers object that the theory postulates unobservable entities; others believe that it is self-contradictory to speak of unconscious mental processes, for what is mental is, by definition, conscious.

In answer to the first objection, it can be pointed out that this practice is common in the most respectable parts of science. Electromagnetic fields and energy quanta are as unobservable as unconscious fantasy. They are, nonetheless, scientifically legitimate because of the functions performed by the theories embodying them, a point to which we shall return. In answer to the second objection, it may be admitted that psychoanalytic theory involves some stretching of such terms as "desire" and "thought" (as in the unconscious thoughts believed to underlie the conscious content of a dream). But, again, this is standard practice in scientific theorizing. The sub-microscopic particles postulated in the kinetic theory of gases are modeled on familiar physical objects, like baseballs, except that they lack some of the properties of baseballs, like color and texture, and they possess perfect elasticity. One may as well say that it is a contradiction to speak of physical particles that have no color. Difference from familiar concepts is not in itself fatal. Again, the cru-

cial question is what can be done with the concepts thus derived.

The serious difficulties emerge when we try to determine whether psychoanalytic concepts have the kind of status that is required for scientific validity and fruitfulness. This problem has two closely related parts. (1) Do psychoanalytic terms have any empirical significance, and if they do, how can it be exhibited? (2) How can theoretical principles couched in these terms be put to an empirical test? These questions become two sides of the same coin if we make certain assumptions that are widely shared by contemporary philosophers of science. First, a term has the kind of semantic status required for science if and only if statements in which it figures have implications for what would be experienced under certain circumstances. Second, one brings out a term's empirical or scientific significance, as contrasted with its pictorial associations, by tracing out such implications. Third, it is only if statements have such implications that they can be put to an empirical test. Given these assumptions, we can deal with the two questions simultaneously. By showing how statements involving the term *repress* give rise to implications of a sort that make an empirical test possible, we will at the same time be showing what scientific significance the term has over and above any of its pictorial associations—for example, a man firmly clamping a lid down on a pot of molten metal. With this equivalence in mind, the following discussion will be explicitly directed to the second question: How can the theoretical principles of psychoanalysis be empirically tested?

There is a commonly accepted doctrine, largely derived from a consideration of physics, according to which a theory involving unobservables gets empirical significance by virtue of the fact that it, together with subsidiary assumptions, implies various general lawlike hypotheses that can be directly tested empirically. In this way the theory can be assessed in terms of the extent to which it succeeds in explaining and unifying a variety of lower-level laws that have been empirically confirmed and, on the negative side, the extent to which it does not imply lower-level hypotheses that have been empirically disconfirmed. The Bohr theory of atomic structure, which represents an atom as a sort of miniature solar system with electrons revolving in orbits around the nucleus, cannot be tested directly, for an individual atom cannot be observed. However, from the theory we can derive a variety of testable hypotheses—for instance, those concerning the constitution of the spectrum of the light emitted from a given element.

## DERIVING TREATABLE HYPOTHESES

One might well expect to have difficulty deriving testable hypotheses from psychoanalytic theory. The theory represents the postulated unconscious processes mediating between events that are accessible to either introspection or observation, just as do unobservable processes within the atom in the Bohr theory.

In a typical sequence we start with conscious Oedipal desires in a child. Tentative attempts at satisfaction of the desires are met with violent opposition, and as a result the child builds up strong fear and/or horror of the realization of the Oedipal desires. Thus far, everything is, in principle, directly accessible to one or more observers. Then, according to the theory, the complex of desires, fears, and guilt is repressed, whereupon it undergoes various transformations, the exact nature of which is influenced by things that happen to the person, these things again being directly observable. In particular, the associations formed in the unconscious are largely determined by conscious experiences of the person. Finally, the unconscious complex is manifested in various ways—dreams, memory failures, slips of the tongue, compulsions, obsessions, psychosomatic illnesses—all of which are again accessible to experience. This being the case, one would suppose that the theory would yield general hypotheses to the effect that whenever strong desires of a certain kind are met with strong internal and/or external opposition, then (perhaps with the further assumption of certain kinds of intervening experiences) abnormal symptoms of certain kinds will be forthcoming. In other words, since unconscious psychic processes are supposed to provide connecting links between observables, a theory about them should imply that certain antecedent observables would lead to certain consequent observables.

In fact, however, we find little of this. Some attempts have been made to derive hypotheses about statistical distributions from parts of the theory. For example, the theory of dreams holds that dreams partially satisfy repressed desires by representing them as satisfied. It would follow from this that if a group of people were prevented from dreaming for several nights, they would then show a higher average level of tension than a control group. This hypothesis has been tested, using eyeball movement as a criterion of the occurrence of dreams. Most efforts of this sort have stemmed from relatively peripheral components of the theory; in particular, virtually nothing has been done to derive testable hypotheses specifying sufficient conditions for the occurrence of abnormal symptoms. It is only if this were done that the theory could be used for the prediction of such phenom-

ena. Perhaps this is because of the psychoanalyst's preoccupation with the treatment of particular cases rather than with controlled testing of general hypotheses.

There are other features of the situation that also make the formulation of testable hypotheses extraordinarily difficult. Psychoanalytic theory has not been developed to the point where one can give sufficient conditions for one outcome rather than another even on the theoretical level of unconscious processes. Repression is said to occur when a desire arouses great anxiety, but just how much anxiety is required? Obviously, the amount is crucial, but the measurement problem has yet to be solved. Again, given a certain level of anxiety aroused by Oedipal desires, repression is not the only possible outcome. There might, instead, be a regression to the oral or anal phase, or the libido might be redirected into homosexual channels. There are some suggestions about what makes the difference—for example, if one never fully outgrew an earlier stage, this makes regression more likely. But at present this is all rather loose.

Moreover, once repression has occurred, the repressed material may develop in a great many different ways. The fear of sexual contact with the mother may or may not generalize, and if it does, it may generalize along various dimensions. Thus, the person may develop a dread of sexual contact with anyone or only with anyone who is like his mother in some respect. A part of the complex may come to be associated with things that have little or no intrinsic connection with it, as the girl in the example cited above formed an association between sexual intercourse and the breaking of a vase. It may well seem impossible to develop principles that would take into account all the determinants of unconscious trains of thought in a way that makes possible, in principle, the prediction of such associations. This impression is reinforced by the fact that these associations are often powerfully influenced by the person's external experiences, which could not be predicted on the basis of psychological facts about him. Thus, in the above example the girl had once broken a vase and cut her finger, which had bled profusely, an incident that then was associated in her mind with the bleeding accompanying defloration.

But even if connections were strong on the level of unconscious processes, there would still remain the job of formulating sufficient conditions for the occurrence of the ultimate facts to be explained. One and the same unconscious complex, given our present powers of discrimination, may issue in a phobia, hysterical paralysis or anesthesia, obsessive concern over bodily symptoms, or a generalized feeling of unworthiness, to mention only a

few possibilities. No doubt the choice of symptom is due to other factors, but the problem has not been investigated sufficiently to yield even promising general hypotheses.

## BACKGROUND FOR CLINICAL INTERPRETATION

In view of the extreme difficulty of empirically verifying psychoanalytic theory, one might ask why it should be regarded as anything other than an imaginatively satisfying fantasy. Why does it seem to have an empirical foundation? The answer is that it has significant connections with empirical facts but not connections of the sort insisted on by philosophers of science who take their models from physical theory. Psychoanalytic theory has grown out of the clinical treatment of neurotics, and in that context it has the function of providing suggestions for the interpretation of particular cases. Thus, if we are dealing with a compulsion neurosis, the theory tells us that compulsive behavior simultaneously provides substitute satisfactions for repressed desires (through the realization of states of affairs unconsciously associated with the realization of the desires) and guards against the arousal and/or satisfaction of the desire. (See the clinical case described above.)

Furthermore, the theory tells us what kinds of desires are most often repressed—incestuous, homosexual, aggressive. Also, psychoanalytic theory is associated with certain techniques—the analysis of dreams, of free associations, and of reactions to the analyst—for ferreting out repressed material in particular cases. Thus, the theory provides leads for the analyst. Insofar as it has this function rather than that of explaining and unifying testable hypotheses about the conditions under which, in general, we will get one outcome rather than another, it is no defect that it is largely made up of rather loose statements about what can happen, given certain conditions, and what can be responsible for a given symptom. In explaining an event, *E*, that has already occurred, our needs are simpler than when we are engaged in predicting or establishing general principles. In retrospective explanation we can take advantage of our knowledge that *E* has already occurred; we are reasoning backward to its sources. Therefore, provided we have a list of possible causes and some way of telling which of these are present, we have something to go on, even if each statement of possible cause is only to the effect that *C* can result in *E*. If we were setting out to predict, however, we would need a further specification of the conditions under which *C* will in fact lead to *E*. The knowledge that an unconscious desire for

and fear of intercourse with the father, plus an association between intercourse and breaking a vase, can lead to a compulsive tendency to arrange vases so as to minimize chances of breakage is general knowledge of a sort, but not of the sort exemplified by the Newtonian theory of gravitation, in which the general principles enable one to predict one state of the system from any other state of the system.

Thus, one can say that psychoanalytic theory, given the way it has developed up to now, makes contact with empirical reality through being used as a basis for explanations of certain kinds of observable occurrences and that the theory receives empirical support to the extent that such explanations are adequate. To many methodologists this situation is profoundly unsatisfying. If a theory yields predictively confirmed hypotheses, we have a strong indication that contact with something real has been made, for by thinking in these terms, we have succeeded in anticipating the course of nature. But if the theory can provide only suggestions for retrospective explanations, it is not so clear what this shows. More specifically, many have suspected that the success of psychoanalysts in devising explanations of their patients' symptoms is more a function of the analysts' ingenuity than of the soundness of their theory. It is easy to get the impression that a plausible explanation in psychoanalytic terms could be framed for any behavior, no matter what the facts. If it is not a reaction formation from overattachment to mother, then it is a projection of a self-directed death wish, and so on.

## ADEQUACY OF CLINICAL INTERPRETATIONS

Clearly, what is needed is a set of objective criteria for the adequacy of an explanation in terms of unconscious psychic factors, criteria that would permit us to assess a proposed explanation on some grounds other than the way it seems to make sense of the phenomena. If and only if such criteria can be formulated can explanations of particular cases provide any empirical basis for the theory.

Within the limits of this article, we can only touch briefly on the problems involved in formulating and defending such criteria. The problems fall into three groups.

**STATUS OF THE DATA.** Questions have often been raised about the status of the ultimate data to which the psychoanalyst appeals in justifying an interpretation. These consist of the behavior of the patient, verbal and

otherwise, in therapeutic sessions. Criticisms have been of three sorts.

First, the data actually presented are a small sample of all the behavior engaged in by the patient in the presence of the analyst. We are almost never given any reason for supposing that this is a representative sample, that the analyst has not, perhaps unconsciously, selected those items that best support his hypothesis.

Second, a given patient is rarely, if ever, compared with controls who do not have his difficulties. Without this we cannot show that the data cited have bearing on the abnormalities to be explained. For example, if almost anyone would get annoyed when the analyst acts bored with the session, then the fact that patient *A* does so is not likely to reveal anything that is responsible for any idiosyncrasy of his.

Third, the analyst may often be guilty of contaminating the data through, perhaps unconsciously, tipping the patient off about his interpretation, thus implicitly inviting the patient to produce associations that will support that interpretation.

These are serious problems in data collection and assessment, and they will have to be solved if psychoanalysis is to become more respectable scientifically. But since it seems in principle possible to overcome them, they are less crucial for the logical status of the theory than problems in the other groups.

**UNCONSCIOUS CAUSES.** An explanation of *E* in terms of *C* is not warranted unless *C* actually exists. What objective tests are there for the actual existence of the unconscious psychic factors appealed to by the analyst? Analysts regularly use a number of detection procedures.

Among the things they consider significant are the following: (1) Patterns of behavior that are as they would be if *A* had a desire of which he is not conscious. For example, a seventeen-year-old girl devotes a great deal of time and energy to the small children of a youngish widower friend of the family, though she is not aware of being in love with him. (2) Patterns of feeling that have the same status. In the same example, the girl gets very depressed when the widower does not send her a birthday present. (3) Analysis of dreams and of free associations. Such analysis proceeds in a rather devious fashion and cannot be illustrated briefly. It is based on the principle that unconscious complexes influence conscious thought and fantasy, including dreaming, by producing relatively safe conscious derivatives of these complexes. (4) Final realization by the patient, after treatment, that he had the desire in question all along.

The inferences involved in the use of these procedures are extremely complex, and it is difficult to say just how conclusively anyone has ever demonstrated the existence of certain unconscious material in a given case. It is worth noting that the use of (3) and (4), unlike (1) and (2), requires the assumption of certain parts of the theory. Thus, for example, we cannot take dreams to reveal unconscious desires in the way analysts do unless we assume that dreams are formed in the manner postulated by the theory. This means that insofar as explanations that are supported in part by dream interpretation are adduced in support of the theory, we are going round in a circle.

**UNCONSCIOUS COMPLEXES AND SYMPTOMS.** The most difficult problem is that of showing that a given unconscious complex is responsible for certain symptoms. Granted that the girl does have a repressed desire for and dread of sexual intercourse with her father, why should we suppose that this is what led her to develop a compulsive tendency to arrange the vases in her room in a certain way before retiring? In order to answer this question, we shall have to decide what kind of explanation this is supposed to be. Freud often gives the impression that it has the ordinary pattern of an “in-order-to” explanation (“I went into the kitchen in order to get a bottle of beer” or “I went into the kitchen because I wanted a bottle of beer”), except that here the want is unconscious. But the ordinary “in-order-to” explanation carries the assumption that the agent believes that the action in question is, or may be, instrumental in the satisfaction of the want in question. Can we say that the girl unconsciously believed that preventing the vases from breaking would be instrumental in preventing intercourse with her father? A strange belief, but Freud did say that the unconscious is quite illogical. Or should we say, rather, that no belief is involved here but only an association between breaking a vase and intercourse? However this issue is resolved, this assimilation will not help us to justify the explanation, for the fundamental method of justifying an ordinary “in-order-to” explanation—getting a sincere report by the agent of why he did what he did—is not available here.

Freud might claim that an analogue is available—the realization by the patient, after treatment, that that was why she had to arrange the vases as she did. However, if one rests the adequacy of the explanation on the patient’s posttherapeutic insight, he leaves himself open to the charge of undue influence on the source of data. Moreover, circularity comes up again, for if the patient came to have this conviction as a result of being presented with

this explanation under hypnosis, this would not count in favor of the explanation. Only insight that comes after certain kinds of therapeutic interactions is relevant, and the claim that insight produced in that way is valid depends on the psychoanalytic theory about the effects that can be expected from psychoanalytic therapy. Thus, there are difficulties in construing the explanation on the model of "I went to the kitchen because I wanted a bottle of beer." On the other hand, if we take as our model an everyday explanation in terms of physical causation, like "The window broke because a baseball hit it," we will have to support it by reference to general principles to the effect that factors of the sort cited have results of the kind we are seeking to explain. And the absence of such tested generalizations in psychoanalysis has already been noted.

Thus, it would seem that before psychoanalytic theory can enjoy a firm empirical foundation, its practitioners must either develop explicit and workable objective criteria for the adequacy of interpretations of clinical phenomena in terms of unconscious factors, or do more to derive testable general hypotheses from the theory, or do both.

**See also** Dreams; Existential Psychoanalysis; Freud, Sigmund; Psychoanalysis; Psychology; Religion, Psychological Explanations of; Unconscious.

### Bibliography

The basic expositions of psychoanalytic theory are to be found in the works of Sigmund Freud. See especially his *General Introduction to Psychoanalysis* (New York: Garden City, 1949), *New Introductory Lectures on Psychoanalysis* (New York: Norton, 1933), *The Interpretation of Dreams* (New York: Macmillan, 1937), *The Ego and the Id* (London: Hogarth Press, 1927), *Inhibitions, Symptoms, and Anxiety* (London, 1961), and the theoretical essays in Vol. IV of the *Collected Papers* (London: Hogarth Press, 1948).

Important later additions include Anna Freud, *The Ego and the Mechanisms of Defence* (London: L. and Virginia Woolf at the Hogarth Press, and the Institute of Psycho-analysis, 1937), and the new developments in ego psychology to be found in two articles by Heinz Hartmann, "Comments on the Psychoanalytic Theory of the Ego," in *The Psychoanalytic Study of the Child*, Vol. V (New Haven, CT: Yale University Press, 1950), and "The Mutual Influences in the Development of the Ego and Id," *ibid.*, Vol. VII (New Haven, CT: Yale University Press, 1952).

Useful expositions are to be found in R. L. Munroe, *Schools of Psychoanalytic Thought* (New York: Dryden, 1955); Otto Fenichel, *The Psychoanalytic Theory of Neurosis* (New York: Norton, 1945); C. S. Hall, *A Primer of Freudian Psychology* (Cleveland, OH: World Publishing, 1954); and David Rapaport, "The Structure of Psychoanalytic Theory," in *Psychology: A Study of a Science*, edited by Sigmund Koch, Vol. III (New York: McGraw-Hill, 1959).

Methodological problems are explored in *Psychoanalysis, Scientific Method, and Philosophy*, edited by Sidney Hook (New York: New York University Press, 1959); B. F. Skinner, "Critique of Psychoanalytic Concepts and Theories," in *Minnesota Studies in the Philosophy of Science*, Vol. I (Minneapolis: University of Minnesota Press, 1956); Antony Flew, "Motives and the Unconscious," *ibid.*; Gustav Bergmann, "Psychoanalysis and Experimental Psychology," in *Psychological Theory*, edited by M. H. Marx (New York: Macmillan, 1951); Ch. 3 of R. S. Peters, *The Concept of Motivation* (London: Routledge and K. Paul, 1958); a series of essays reprinted in *Philosophy and Analysis*, edited by Margaret Macdonald (Oxford, 1954)—Stephen Toulmin's "The Logical Status of Psycho-analysis," Antony Flew's "Psycho-analytic Explanations," and R. S. Peters's "Cause, Cure, and Motive"; and B. A. Farrell's articles "The Criteria for a Psychoanalytic Interpretation," in *PAS*, *supp.*, 36 (1962)—reprinted in *Essays in Philosophical Psychology*, edited by D. F. Gustafson (Garden City, NY: Anchor, 1964), pp. 299–323—and "Can Psychoanalysis Be Refuted?" in *Inquiry* 4 (1961): 16–36.

Attempts to derive testable general hypotheses from psychoanalytic theory are canvassed in R. R. Sears, *Survey of Objective Studies of Psychoanalytic Concepts* (New York: Social Science Research Council, 1943).

William P. Alston (1967)

## PSYCHOKINESIS

See *Parapsychology*

## PSYCHOLOGISM

"Psychologism" is the term first used in Germany in the first half of the nineteenth century to designate the philosophical trend defended by Jakob Friedrich Fries (1773–1843) and by Friedrich Eduard Beneke (1798–1854) against the dominant Hegelianism. Fries and Beneke advocated a philosophical position based entirely on psychology. They held that the only instrument philosophical inquiry has at its disposal is self-observation (or introspection) and that there is no way to establish any truth other than by reducing it to the subjective elements of self-observation. Psychology becomes, from this point of view, the fundamental philosophical discipline. Logic, ethics, metaphysics, philosophy of law, philosophy of religion, and philosophy of education are all little more than psychology or applied psychology. Beneke wrote, "With all of the concepts of the philosophical disciplines, only what is formed in the human soul according to the laws of its development can be thought; if these laws are understood with certainty and clarity, then a certain and clear knowledge of those disciplines is

likewise achieved" (*Die Philosophie in ihrem Verhältnis zur Erfahrung*, p. xv).

Fries and Beneke, who viewed Immanuel Kant as their predecessor inasmuch as he defended the "rights" of experience, held, nevertheless, that he was mistaken in wanting to institute an inquiry independent of experience which would arrive at knowledge of the a priori forms of intuition and of the categories and in seeking the transcendental ground of truth—the objective validity of human knowledge. This inquiry, Fries claimed, is impossible. The critique of reason can only be a science of experience based on self-observation (*System der Metaphysik*, p. 110). In the same period Vincenzo Gioberti branded as psychologism all of modern philosophy from René Descartes on. He meant by psychologism the philosophical procedure that claimed to go from man (that is, from experience) to God and contrasted it with ontologism, which is the movement from God to man.

The doctrine defended by Fries and Beneke has some connection with certain aspects of English empiricism from John Locke to David Hume in that in both theories experience is not only the instrument of control and the criterion of the truth of knowledge but also the psychological origin of knowledge itself.

Fries and Beneke were correct in accusing Kant of rejecting psychologism, since he had posited the premises for a critique of any psychologism by distinguishing (in a famous passage in the *Critique of Pure Reason*) the *quaestio facti* of the "physiological derivation" of a priori concepts—that is, of their occurrence in the mind or consciousness of man—from the *quaestio juris* of their validity, which demands as a response the transcendental deduction. This distinction, on the basis of which Kant criticized Locke, who would have answered only the first question, is one of the pivotal points of the whole Kantian doctrine—namely, that the truth of empirical knowledge does not depend on the psychological mechanism but on a priori conditions independent of this mechanism; that the validity of the moral norm does not depend on desires or appetites but is a priori as well; and that the validity of aesthetic judgments is in turn based on taste, an a priori faculty.

Toward the middle of the nineteenth century, psychologism was defended in the very field in which it would seem most foreign—logic and mathematics. In John Stuart Mill's *A System of Logic* it is explicitly stated that introspection is the only basis of the axioms of mathematics and the principles of logic; in Mill's *Examination of Sir William Hamilton's Philosophy* logic is classified under psychology and distinguished from it only as the

part is distinguished from the whole or art from science. Many logicians in subsequent years accepted this point of view.

The Kantian point of view was developed systematically by Rudolf Hermann Lotze in his *Logik*. The psychological act of thinking is, according to Lotze, completely distinct from the content of thought. The psychological act exists only as a determinate temporal phenomenon, whereas the content has another mode of being—validity. A decade later Gottlob Frege defended the same point of view with regard to mathematics.

Never take a description of the origin of an idea for a definition, or an account of the mental and physical conditions through which we become conscious of a proposition for a proof of it. A proposition may be thought, and again it may be true; never confuse these two things. We must remind ourselves, it seems, that a proposition no more ceases to be true when I cease to think of it than the sun ceases to exist when I shut my eyes. (*Die Grundlagen der Arithmetik*, introduction)

In the last decades of the nineteenth century, the neo-Kantians argued against the psychologistic presentation of philosophy. The Baden school (Wilhelm Windelband, Heinrich Rickert) defended the independence of values from psychological experience, which could never establish their absoluteness and necessity, and the Marburg school (Hermann Cohen, Paul Natorp) held, similarly, that the validity of science, like that of ethics and aesthetics, does not depend on psychological conditions but on the laws proper to these sciences—that is, on the methodological rules that govern their construction. Cohen and Natorp held, moreover, that "thought" or "consciousness" does not designate a psychic reality subject to introspection but the objectively valid content of knowledge—the totality of the possible objects of knowledge itself and the method used in the development of the sciences.

The systematic critique of psychologism in the fields of logic and mathematics is an important part of Edmund Husserl's *Logische Untersuchungen*. His main objections are that if logical laws were based on psychological laws, then (1) they ought to be, like the latter, vague and approximate, whereas, at least in part, they are so exact that they cannot be guaranteed by an empirical element; (2) they ought to be based, like all empirical laws, on induction, which yields only a probable validity and not the apodictic certainty they manifest; (3) they ought to imply the existence of such psychic events as representation and judgment, whereas they do not con-

cern the reality of psychic life and of other facts (unlike the laws of nature, which are merely probable) but concern necessary relations independently of facts (*Logische Untersuchungen*, Vol. I, Secs. 21–24). Later in his career Husserl wrote, in terms very close to Frege's, "To refer to it [a number] as a mental construct is an absurdity, an offence against the perfectly clear meaning of arithmetic discourse, which can at any time be perceived as valid, and precedes all theories concerning it" (*Ideen*, Sec. 22). He warned against the tendency to "psychologize the eidetic"—that is, to identify essences, which are the authentic objects of knowledge, with the simultaneous consciousness of these essences (*ibid.*, Sec. 61).

The battle between psychologism and antipsychologism is sometimes fought among philosophers with the same point of view. Among the existentialists Martin Heidegger, who adopted as his method Husserl's phenomenology, intended existential analysis as the uncovering of human situations in their essence, not in their psychic occurrence (*Sein und Zeit*, Halle, 1927, Sec. 7), whereas Jean-Paul Sartre, speaking of existential psychoanalysis, seems inclined toward psychologism, although he tried to correct it by affirming that "consciousness is not a mode of particular knowledge but it is the dimension of transphenomenal being in the subject" (*L'être et le néant*, Paris, 1943, p. 17).

Within logical empiricism the argument against psychologism is one of the fundamental points of Rudolf Carnap's first work, *Der logische Aufbau der Welt*. The fundamental theses of *Logische Syntax der Sprache*, especially the principle of tolerance, are incompatible with psychologism, according to which, obviously, there could be only a single language—that determined by psychological laws. Carnap took the same line when he criticized Bertrand Russell's thesis that propositions are mental events in "Empiricism, Semantics, and Ontology." Arguments against psychologism occur frequently in the writings of other logical empiricists, though traces of psychologism can be found in the thesis, deriving from Russell and held by many logical empiricists, of the immediate, private, and incommunicable character of the sense data that are at the basis of empirical propositions.

**See also** Beneke, Friedrich Eduard; Carnap, Rudolf; Cohen, Hermann; Descartes, René; Empiricism; Existential Psychoanalysis; Frege, Gottlob; Fries, Jakob Friedrich; Gioberti, Vincenzo; Hegelianism; Heidegger, Martin; Hume, David; Husserl, Edmund; Intuition; Kant, Immanuel; Locke, John; Logical Positivism; Lotze, Rudolf Hermann; Mill, John Stuart; Natorp, Paul; Neo-Kantianism; Propositions; Psychology; Rick-

ert, Heinrich; Russell, Bertrand Arthur William; Sartre, Jean-Paul; Windelband, Wilhelm.

## Bibliography

### WORKS FAVORING PSYCHOLOGISM

- Beneke, Friedrich Eduard. *Die Philosophie in ihrem Verhältnis zur Erfahrung, zur Spekulation, und zum Leben*. Berlin, 1833.
- Fries, Jakob Friedrich. *System der Metaphysik*. Heidelberg, 1824.
- Gioberti, Vincenzo. *Introduzione allo studio della filosofia*. Brussels, 1840.
- Lipps, Theodor. *Grundzüge der Logik*. Hamburg–Leipzig, 1893.
- Lotze, Hermann. *Logik*. Leipzig, 1874. Translated by Helen Dendy as *Logic*. Oxford, 1884.
- Mill, John Stuart. *Examination of Sir William Hamilton's Philosophy*. London, 1865.
- Mill, John Stuart. *A System of Logic*. 2 vols. London, 1843.
- Sigwart, Christoff. *Logik*. 2 vols. Tübingen, 1873–1878. Translated by Helen Dendy as *Logic*. 2 vols. London, 1890.

### WORKS CRITICAL OF PSYCHOLOGISM

- Carnap, Rudolf. "Empiricism, Semantics, and Ontology." *Revue internationale de philosophie*, 4th year (11) (1950): 20–40. Reprinted in *Readings in the Philosophy of Science*, edited by Herbert Feigl and May Brodbeck. New York: Appleton-Century-Crofts, 1953.
- Carnap, Rudolf. *Der logische Aufbau der Welt*. Berlin: Weltkreis, 1928.
- Carnap, Rudolf. *Logische Syntax der Sprache*. Vienna: Springer, 1934. Translated by Amethe Smeaton as *The Logical Syntax of Language*. London: Kegan Paul, 1938.
- Frege, Gottlob. *Die Grundlagen der Arithmetik*. Breslau, 1884. Translated by J. L. Austin as *The Foundations of Arithmetic*. Oxford: Blackwell, 1950.
- Husserl, Edmund. *Ideen zu einer reinen Phänomenologie und Phänomenologischen Philosophie*. Vol. I. Halle: Niemeyer, 1913. Translated by W. R. Boyce Gibson as *Ideas: General Introduction to Pure Phenomenology*. New York: Macmillan, 1931.
- Husserl, Edmund. *Logische Untersuchungen*. 2 vols. Halle: Niemeyer, 1900–1901.
- Pap, Arthur. *Elements of Analytic Philosophy*. New York: Macmillan, 1949.

### OTHER RECOMMENDED TITLES

- Bermudez, Jose Luis. "Psychologism and Psychology." *Inquiry* 42 (1999): 487–504.
- Cohen, Jonathan. "Frege and Psychologism." *Philosophical Papers* 27 (1998): 45–67.
- Davis, Wayne. "Psychologism and Humeanism." *Philosophy and Phenomenological Research* 67 (2003): 452–459.
- George, Rolf. "Psychologism in Logic: Bacon to Bolzano." *Philosophy and Rhetoric* 30 (1997): 213–242.
- Kitcher, Patricia. "Revisiting Kant's Epistemology: Skepticism and Psychologism." *Nous* 29 (1995): 285–315.

**Nicola Abbagnano (1967)**

*Translated by Nino Langiulli (1967)*

*Bibliography updated by Benjamin Fiedor (2005)*



## PSYCHOLOGY

In the development of “psychology,” the study of the mental life and activities of animals and men, three phases can be conveniently distinguished—the presystematic, the systematic but prescientific, and the scientific. The presystematic, by far the longest of the three phases, is that in which men observed and reflected on human ways and embodied their reflections in aphorisms, anecdotes, and fables. Presystematic thinking is important since it has been passed down through the ages and is continually augmented by that amalgam of wisdom, superstition, and dogma that those who claim no professional competence like to describe as the fruits of their experience. The presystematic psychology of contemporary primitive groups has been recorded by anthropologists, but little is known of the corresponding ideas of the precursors of the systematic psychology of the European tradition. The doctrines of the pre-Socratic philosophers are transitional.

### SYSTEMATIC PHILOSOPHY OF MIND

**MIND, BODY, AND NATURE.** Systematic psychology began with Aristotle’s *De Anima*, which was of outstanding importance at an early stage because it provided a solid, biologically based conceptual scheme. This involved, first, an elucidation of the concept of soul (*ψύχη*) and such related concepts as mind (*νοῦς*), which were regarded as the differentiating properties of the phenomena to be studied. Aristotle’s scheme laid down the lines along which the relationship between various manifestations of soul and mind were conceived until the seventeenth century.

Second, life and mind, being closely connected with the functioning of the body, must be conceived of in a way that does justice to the peculiar intimacy of this relationship. Aristotle paid close attention to this relationship.

Third, there is the problem of how the relationship between psychological phenomena and other phenomena of the natural world is to be conceived. Are psychological concepts and categories of explanation reducible to others? Aristotle, again, was particularly interested in this question because of the attempts of some of his contemporaries and predecessors to show that human behavior fell under the concept of motion, which had a wide applicability in the natural world.

In the exposition of the systematic period of psychology these problems will be employed not simply as a framework for expounding the main lines of Aristotle’s

system of psychology but also as a framework for picking out the main features of the most important theoretical systems since Aristotle laid the foundation of psychology.

**PLATO AND ARISTOTLE.** Aristotle (384–322 BCE) insisted on the widest possible definition of soul, thus returning to the pre-Platonic view that soul is virtually the principal of all life. The natural expression for a living thing was *ἐμψυχον σῶμα*—“body with a soul.” Aristotle started from the linguistic point that some bodies are so described whereas others are not and asked by what criterion this distinction was made. His answer was that it is life but that there are different levels of life. Intellect, sensation, nutrition, motion, are all forms of being alive. What they have in common, however, is a self-originating tendency to persist toward an end.

This marked both a return to and a great improvement on pre-Platonic views of soul. In early Greek thought soul was thought of simply as that which keeps a man alive and which leaves his body when he dies. It was connected with breathing. Spirit (*θύμος*), on the other hand, was thought of as the generator of movement; it was connected with the movement of the limbs and with emotional states. It was thought of as quite distinct both from soul and from mind, which was regarded as the source of images and ideas. The notion of the soul as a whole of which spirit and mind were attributes emerged only gradually.

Plato (427?–347) tried to combine the concept of the soul as a whole with a stress on the preeminence of mind, which he inherited from Anaxagoras. His account, therefore, of the soul as a whole was constantly confused by the special status that he accorded to mind. In the *Republic* he spoke of the soul as having three parts—reason or mind, spirit, and desire (*ἐπιθυμία*). But he also thought that reason was the defining property of an immaterial substance that survived bodily death whereas spirit and desire passed away with the body. Similarly, in the cognitive sphere he regarded sensation and imagination as inferior to reason and as intimately connected with the body. This represented a fusion of the Orphic belief in the survival of the soul with an exaltation of mathematical reasoning as the only way of obtaining certain knowledge, which Plato took from the Pythagoreans. He thought that in mathematics the soul grasps forms that are eternal and nondeceptive. As like can be known only by like, the soul, in its rational aspect, must also be eternal. Plato’s conviction was reinforced by such considerations as those that he adduced in the *Meno*, in which the grasp of mathematical truths was exhibited in an untutored slave. According to

Plato, this indicated that the slave was being made to remember what he had known previous to his embodiment. Thus, Plato's preoccupation with epistemology led him to make a sharp cleavage between the rational and irrational parts of the soul.

Aristotle approached the matter from a biological rather than an epistemological standpoint. Reason, spirit, and desire represented different levels of being alive. To be alive is to possess a self-originating tendency toward an end. This is exhibited at the lowest level in nutrition and reproduction. Thus, plants have a low-grade soul. Animals have sensation, locomotion, and desire superimposed upon nutrition and reproduction. Human beings, in addition, have reason, or mind, by means of which a rule or plan is imposed upon desire. By mind is meant self-direction in accordance with a rational formula.

Aristotle maintained that the lower level of soul is a necessary condition for the higher and that the possession of a higher type of soul also changes the way in which the lower functions. Because humans are rational, they feed, reproduce, perceive, and act in a manner that differs from that of animals.

*Soul and body.* Plato's view of the special status of reason was plausible at a time when almost nothing was known about the functioning of the brain and nervous system, for abstract thought seems to proceed with little dependence on bodily organs. Furthermore, the identity of a subject of experience through time does not seem to depend entirely on bodily continuity. There is thus a case for Plato's concept of the rational soul as some kind of active agency that inhabits the body for a brief period.

Plato thought that the rational soul inhabits the head because the head is round (the most perfect shape and, hence, an appropriate place for the seat of reason) and the part of the body nearest the heavens. It makes contact with the brain, which was conceived of as a kind of marrow encased in the skull. The irrational soul makes contact with the marrow of the spinal cord in its bony sheath. The better part of the irrational soul, spirit, inhabits the heart and functions in such manifestations of life as energy, courage, and ambition; the worse part, desire, functions below the diaphragm, in appetite, nutrition, and reproduction. The rational and irrational parts affect each other through the liver, which acts as a sort of mirror of thought.

In sleep the soul is shut up, and its motions subside. A few agitations remain, however, and produce dreams. Usually dreams are the expressions of desires that are suppressed—an interesting anticipation of Sigmund

Freud's theory of dreams. The good man controls his desires sensibly and so is not unduly disturbed by them in sleep. In the *Republic* Plato also suggested that in sleep the rational soul, if not troubled by irrational desires, can attain truths not otherwise revealed.

Plato thought of sensation as a transmission of motions. The human body receives an impression from without and responds with an inner motion. Some parts of the body—for instance, the hair and the nails—are subject to shock but do not respond with inner movements. Sense organs, however, are good conductors of motion. Thus, hearing, for instance, is the end product of a kind of shock. By means of air in the cavities of the body a blow is transmitted through the ears to the blood and brain and then to the soul. Knowledge does not consist just in sensation but in the activity of the soul in relation to what is thus transmitted. This transmission is complicated by the intervention of memory, imagination, feeling, and association, all of which act as intermediaries between reason and sensation.

Aristotle believed that there was a very intimate connection between soul and body that was a particular case of the more general relationship between form and matter. The soul is "the first actuality of a natural body furnished with organs." He used other examples to illustrate this relationship. If the eye were an animal, he said, eyesight would be its soul, this being the form or capacity of the eye. To speak of soul is to speak of a capacity or propensity to function in a certain way that depends on a certain bodily structure, or it is to speak of the actual exercise of such a capacity or propensity, which is the second kind of actuality. Thus, anger, for instance, can be the appetite of returning pain for pain or the boiling of the blood around the heart, depending on whether the dialectician or the physical scientist is considering it; there is always a biological and a psychological account to be given.

The soul, Aristotle argued, is the cause of the body in three ways. It is its efficient cause in that reference to some concept, such as desire, is required to explain movement. It is the formal cause in that behavior is explained as the exercise of a capacity or tendency. It is the final cause in that reference must be made to "the reason for the sake of which" movements of the body take place. If the behavior is explained by recourse to the rational soul, then plans and rules are imposed on desire. In choice, for instance, means are worked out and adapted to attain an end.

Generally speaking, Aristotle held that soul and body are a particular case of the more general correlatives,

form and matter. When he spoke of theoretical reason rather than practical reason, he suggested that the distinction between matter and form is again exemplified in that reason is both passive and active. But he hinted at another sort of doctrine when he also claimed that active reason comes from without and is divine. It is like a helmsman in a ship. This looks like a concession to the Platonic view of mind.

The details of Aristotle's physiology were carefully related to his idea of the levels of soul. The primary function of the nutritive soul is the absorption of nourishment, but its end is to generate another being like itself. The unity of the species is thus preserved though individual members perish. The stomach was thought of as an oven where animal heat cooks the food and blood in the heart. The heart is the seat of life, sensation, motion, and heat.

Sensation is a discriminative power from which the higher cognitive functions develop. There is the organ, the power to receive sensible forms, and the sense, regarded as constituted of both matter and form. In sensation the sense organ is assimilated to its object—for example, the eye becomes colored. But whereas in nutrition both matter and form of external objects are absorbed, in sensation only form without matter is taken in, like wax taking the imprint of a seal ring. Each sense is sensitive to one or more qualities ranging between extremes. Too little would not register; too much would destroy the organ. This was an application of Aristotle's doctrine of the mean that he developed in relation to moral conduct.

The particular senses are all developments of touch, depending on the intervention of a more refined medium. Taste, for instance, apprehends the savory properties of bodies through the intermediary of moisture; smell, the odorous properties conveyed through the air. In the transmission of sensations to the heart and in the vitality that flows from the heart, the "connatural spirits" play an important role. They were thought of as a kind of inner air quite distinct from the outer air that we breathe. Closely associated with the blood, they acted as a universal internal medium for the transmission of sensation. Besides the specific senses there is *sensus communis*, which is not a sixth sense but a generic power of sensation as such which provides unity for the sensitive soul in its particular manifestations. The ear does not see; however, the man who hears also sees, and some qualities are presented through more than one sense—for example, roundness by sight and touch. By *sensus communis* we also perceive the common sensibles of figure, motion,

rest, magnitude, and also what Aristotle called the accidental sensibles, which are the principles of association of ideas—similarity, contiguity, and the like. We also perceive that we perceive through *sensus communis*.

Imagination is a by-product of sensation. Forms provided by sensation are manipulated in the absence of physical objects. Memory is a combination of imagination and *sensus communis*. There is an image of something plus an awareness of its pastness. Recollection is rather different, for it involves the exciting of an image and the release of a whole chain of images joined by habit according to the principles of association. Imagination also provides a link between knowledge and action, for desire presupposes the imagination of an end to be attained. It may be deliberative, if influenced by reason, or merely sensitive. Desire is thus dependent on sensation and thought. In this way Aristotle was able to maintain his three levels of soul by making desire appear at two levels, depending on whether it is rational or irrational.

*Psychological and mechanical concepts.* Aristotle believed not only that there were certain very general concepts, such as form, matter, and change, which could be applied to everything; he also extended teleological categories of explanation—his ill-fated final causes—to all nature. Nature, he thought, was composed of natural kinds that could be classified by genus and differentia, which all had a natural place, and which all tended toward the realization of their essence. "Nature, like mind, always does whatever it does for the sake of something, which something is its end." Such modes of explanation proved singularly unfruitful when extended to the physical world. But because they were taken from the realm of life, where Aristotle, a marine biologist and the son of a doctor, was particularly acute, they fitted very well, in a general sort of way, that realm of phenomena in which they had their natural home. Aristotle was often accused by later mechanists of being anthropomorphic, but there is not much wrong with being anthropomorphic about men. Indeed, those who later attempted to explain human behavior in mechanical terms applicable to the physical world may well have made the obverse mistake to Aristotle's.

Aristotle himself, in criticizing the mechanists of his day, gave some very interesting arguments to show why the soul, which is the source of movement, cannot itself be moved. Plato had steadfastly claimed that the soul was the source of motion. In a famous passage in the *Phaedo* (98B–99D) he made clear his objection to extending mechanical explanations to cover human conduct. Plato admitted that some kind of physical account could be

given of the movements that led up to Socrates' sitting in his prison cell, awaiting his death. But he scorned the suggestion that this account would be a satisfactory explanation of the situation, for an explanation must include some reference to Socrates' reasons for being there. Plato did not, however, develop elaborate arguments against mechanical theories.

Aristotle, on the other hand, wrote his *De Anima* as part of his systematic attempt to classify the different sciences on the basis of the subject matter with which they were concerned. He was therefore very much concerned both with demarcating the field of application of various families of concepts and with sketching the ways in which they were related to each other. Movement (*κίνησις*) was only a particular type of change. He was most anxious to deny that it was either the only or the fundamental type.

Aristotle argued, first, that a logical mistake is made if the soul as a formal cause is thought of as moved in the physical sense. How can a capacity or tendency be conceived of as moving or being moved? Nor can the actualizations of soul in particular cases be properly conceived of as movements, for in practical thought the processes have unity because they go on for the sake of some end. Their particular type of unity cannot be assimilated to such physical unities as the parts of a spatial magnitude; it is more like the unity of a series of numbers. Reference to an end is a conceptual device for picking out how a series of movements are to be thought of as constituting one action; such an end is not itself an extra movement. In the case, too, of some processes of theoretical thought, such as inferring, "thinking has more resemblance to a coming to rest or arrest than to a movement." The end is, as it were, built into the meaning of the term. "Inferring," "concluding," and even "perceiving" are terms that intimate the attainment of ends or standards that are intrinsic to the processes themselves.

**Concept of consciousness.** Arguments of the Aristotelian type have been revived in recent times by such philosophers as Gilbert Ryle, who have defended a predominantly Aristotelian concept of mind in opposition to a Platonic or mechanical concept. Such a concept of mind is in keeping with the biological orientation of psychology that followed the impact of Charles Darwin. However, it sprang out of the post-Wittgenstein reaction against privacy as the hallmark of the mental, which had characterized most psychological theories since the time of René Descartes (1596–1650).

It is difficult for modern Western scholars to grasp that the Greeks really had no concept of consciousness in that they did not class together phenomena as varied as

problem solving, remembering, imagining, perceiving, feeling pain, dreaming, and acting on the grounds that all these are manifestations of being aware or being conscious. Historically, this emphasis on private experience presupposed the development of individualism as a social movement. The Greeks of the city-states lived in a public world of public feats and public concerns. Their word *ιδίωτης*, from which we derive the word *idiot*, was a term of disdain for a man who concerned himself only with private matters. Socrates, with his stress on individual self-knowledge and the care of the individual soul, was a moral innovator. With the conquests of Philip and Alexander the Great and the breakup of the small autonomous Greek states, this moral innovation became systematized in the codes of the Stoics and Epicureans. The ideal of individual self-sufficiency developed as a substitute for the much-lauded self-sufficiency of the city-states. Man, it was claimed, was a citizen of the world who should either discipline himself and purify his individual soul (Stoics) or slip through life unobtrusively by cutting down the possible sources of misery (Epicureans). This led to an increase of interest in the will and the emotions and to an emphasis on individual experience.

This turning inward was institutionalized by Christianity, with its stress on personal salvation and the purity of soul. Introspection vied with revelation as a source of knowledge. St. Augustine paved the way for Descartes's first certainty, *cogito ergo sum*. With Descartes the Platonic view of the soul and of knowledge was reinterpreted in the light of the rise of the mathematical sciences, but there was a difference—the stress on the certainty of our knowledge of our own mental states. Mind was no longer simply associated with reason; it was something to which we have private access and whose rational activity it is self-contradictory to doubt. This stress on privacy as a hallmark of the mental was a far cry from Aristotle's view of soul as characterized by a self-originating tendency to pursue an end. A brief mention, however, should be made of some of the intervening systems, though from the point of view of psychological theory, nothing of any great importance happened after the death of Aristotle in 322 BCE until the seventeenth century, when new systems were inspired by the rise of the physical sciences.

**STOICS AND EPICUREANS.** The Stoics and Epicureans provided an interesting contrast in respect to their views about the relation between soul and the rest of nature. Both attempted a monistic view, but whereas the Stoics reverted to Plato and tried to extend the concept of soul so that it permeated all nature, the Epicureans reverted to

Democritus and extended a mechanical atomistic account of nature to include life and mind.

**Stoics.** The Stoics thought of everything in the universe as being either active or passive; hence, there was no opposition between dead matter and soul. The ultimate substance is fire, which has different forms at different levels of being, ranging from cohesion at the inorganic level, through growth at the plant level, to life of a rational or irrational type at the animal and human level. Fire is thus the all-pervading principle of activity as well as the reason or regulator of change in the universe. Mental activity as found in men is a concentrated form of the universal reason, creatures being vehicles for the operation of this universal regulation. Hence the Stoic injunction to live according to nature, for in simple instinctive tendencies reason is often manifest in an incorrupted form.

The Stoics believed that the soul of man is a very subtle form of the all-pervasive fire, for the corporeal can be affected only by what is corporeal. The soul is affected by the body; therefore, the soul, too, must be corporeal. It combines heat, mobility, and a high degree of rarefaction. Indeed, it was more or less identified with the “connatural spirits” of Aristotle that course through the body closely associated with the blood, which are transmitted in generation, and which are similar in nature to the warm outer air, which is also essential to life. The breast is the seat of the soul.

Perhaps the most interesting and important contribution of the Stoics to psychology was their application of the Aristotelian categories of activity and passivity, which they thought to be the defining attributes of what is real, to the mind. Mental activity, they held, is characterized by assent (*συγκατάθεσις*), which can be exhibited in perception and memory, as well as in practical and intellectual judgment. This may be justified or erroneous, but truth is natural and error unnatural. When error of a perceptual, intellectual, or practical kind occurs, the explanation is to be sought in the theory of emotions or mental disturbances. Basic to this Stoic account was the notion of impulse, which covered both appetite and aversion and which operates obscurely at the level of sensation as well as at the rational level, when it is transformed into the adoption of ends for action. Emotions are thus unsuccessful attempts at full rational choice. The early Stoics left such failures unexplained; the later Stoics assigned the cause to circumstances and, therefore, to things that are beyond our power. From this came their characteristic emphasis on the assertion of will over adversity, of rational choice over irrational promptings.

**Epicureans.** The main interest of Epicurean psychology was its anticipation of mechanical theories of the seventeenth and subsequent centuries. Everything, Epicurus (341–270 BCE) believed, was constructed from atoms and, therefore, everything, including minds, could be explained in terms of the mechanical laws governing atoms. The soul differs from other atoms in that it is lighter and more mobile; heat is fundamental to its nature, but it is not identical with fire. It permeates the body like a subtle air and gives it life.

Sensations are effects produced in sense organs by effluxes from objects, differences in sensations being explained in terms of differences in external movements and in the configurations of the underlying atoms. Similarly, ideas are caused by atoms striking the subtle matter of the thinking soul. Incoming impressions set up other motions in the mind, making possible judgment, which is a motion of the mind superimposed upon an impression. Error occurs when impressions are accompanied by irrelevant motions of the mind. The motions of the mind can be linked together to form complex ideas by principles of association. Reason is simply the use of general ideas brought about by the fusion of images into composite pictures.

It is difficult to see how notions such as error and truth could be generated by such descriptions of mere movements of atoms. Indeed, Epicurus did nothing to meet Aristotle’s acute criticisms of mechanical descriptions of thought. He did something, however, to meet the charge of fatalism in his notorious doctrine of the swerve of the atom, which was a consequence of the self-motion postulated for all atoms. The power of the mind to incline this way or that constitutes its freedom. People are poised between pain, which is one sort of motion, and pleasure, which is an excessive reaction to pain. Between these two extremes there is an equilibrium, which is more permanently satisfying and which reason can guide men to attain. This he called freedom from disturbance (*ἀταράξια*), which is inseparable from the use of reason.

**THEOLOGICAL PSYCHOLOGY.** The psychology of the Greeks had always been, in varying degrees, subservient to epistemological and ethical concerns. The account of reason, for instance, or the role ascribed to the passions was a graphic way of presenting solutions to problems about knowledge and conduct. But there was also the Greek passion for speculation about the ultimate nature of things, about the One in the many, and about the status of mind in the universe and its relation to the body. With the coming of Christianity, which brought with it

the biblical account of the creation of the world, this radical metaphysical speculation abated, and the body was seen largely as something that had to be considered as a potent source of temptation. Psychological theory became almost entirely an offshoot of epistemology and ethics, for the supreme purpose of life for thinking men became the knowledge of God and the quest for salvation.

The religious preoccupations of such writers as Plotinus, Clement, and Augustine introduced, of course, a different emphasis into epistemology and ethics. This was manifest before the coming of Christianity in the work of Philo Judaeus (fl. 20 BCE–40 CE), who thought that real knowledge was a possession only of minds that had been so purified that they received divine illumination. Philo was the first systematic thinker to fuse the religious fervor of the Hebrew tradition with a selection from the conceptual schemes of Plato, Aristotle, and the Stoics. Knowledge of God and a divinely sanctioned code of conduct had somehow to be fitted into the speculative schemes of the Greeks. Because neither God nor his purposes are manifest to the senses, increasing importance was attached to inner experience as a way of knowing. Philo even wrote a treatise titled *On Dreams Sent from God*.

This shift of emphasis from the outer world to the inner world is clearly seen in the Neoplatonism of Plotinus (c. 204–270). Plato, like all the Greeks, was supremely interested in action, politics, and the external world. His theory of Forms was, in the main, explanatory—his version of the search of the Greek cosmologists for the One in the many. Even the supreme Form, the Form of the Good, was both the source of the intelligibility of the world and the supreme ideal of action. Plotinus, on the other hand, saw mystical contemplation and absorption in the One as an end in itself. Psychology therefore became harnessed to the exploration and mapping of inner experience. As G. S. Brett remarks in his *History of Psychology*: “In Plotinus, for the first time in its history, psychology becomes the science of the phenomena of consciousness, conceived as self-consciousness” (R. S. Peters, ed., rev. ed., p. 206).

With the adoption of Christianity as the official religion of the Roman Empire a place had to be found for revelation as well as for knowledge found in inner experience. Augustine (354–430) managed to combine these two sources of knowledge. Insofar as there was no revealed doctrine on a matter, he dealt with it within the framework of Platonism penetrated by Christian mysticism. For instance, the growing knowledge of the self and of God was fitted into a Christianized version of Plato’s

doctrine of reminiscence. Questions about the body, on the other hand, were dealt with by an appeal to the Scriptures. So, too, was the origin of the soul, for it was transmitted into the body when God breathed upon Adam. The lasting influence, however, of Augustine’s *Confessions* was the importance attached to introspection and private experience. No man can escape from his own experience; he can obtain knowledge, insofar as he does not rely on revelation, only by working backward to the presuppositions of his experience as a thinking being. In this approach to the mind Augustine anticipated Descartes.

A corrective to this extreme subjectivity was provided by the rediscovery of Aristotle and the meticulous transmission of his texts by Islamic theologians. The adaptation of Aristotle in the service of Christian theology reached its climax in the work of St. Thomas Aquinas (1224?–1274). But using Aristotle as a substructure to support Christian theology was not entirely straightforward. To start with, there was the problem about the status of reason, one of the most debated topics during the Middle Ages. Aristotle’s account of the Active Intellect suffered from notorious obscurities, and there was the worry about its relation to revelation as well. Furthermore, the Islamic school, culminating in Averroes, had tended to favor a mildly pantheistic interpretation of Aristotle’s doctrine of Active Intellect. Averroes held that the reasons of individuals are but fleeting manifestations of universal reason. Thomas rejected this interpretation, completely following his teacher Albert the Great (c. 1193/1206–1280).

Thomas defined intellect as the faculty of comprehension that each individual possesses as an intelligent being. Nevertheless, reason was still regarded, as by Plato and Aristotle, as the mark of man’s difference from animals and as, in some sense, superhuman. It is qualitatively distinct from sensation and any other processes that are intimately connected with the body.

Apart from this query about the status of reason, which was itself a legacy from Aristotle, Thomas tried to stick to the Aristotelian view of the soul as the form of the body. He deliberately rejected the more Platonic theory that a man is a soul using a body. It was not just respect for the authority of Aristotle that influenced Thomas. The fact was that Christianity was committed to the belief in the resurrection of the body. The intimacy of the connection between soul and body postulated by Aristotle was a better foundation for this doctrine than the more Platonic view occasioning that contempt for the body that culminated in the Albigensian heresy that the body had been created by the devil. Thomas followed Aristotle

closely in his account of sensation, *sensus communis*, memory, and imagination. What was lacking was Aristotle's stress on striving toward an end as the defining characteristic of soul. The intuitive certainties of self-consciousness explored by Augustine remained the foundation both of psychology and of epistemology.

Scholasticism has now become a byword for sustained attention to minor questions within a system whose foundations in revelation were not questioned. There is point in such criticisms. Nevertheless, the Schoolmen preserved and spread a tradition of disciplined discussion that is the lifeblood of science and philosophy. Furthermore, in psychology they handed down not only the general outlines of Aristotle's conceptual scheme but also the details of his psychological system.

The great natural philosophers were nurtured in this Aristotelian tradition even though they eventually overthrew it. At Padua, for instance, where Galileo Galilei was trained, there was a flourishing branch of the Averroistic type of Aristotelianism. Descartes was trained by the Schoolmen at La Flèche, and his *Passions of the Soul* bears witness to these early influences. Even Thomas Hobbes, one of the archenemies of Aristotelian essences, relied on Aristotle's *Rhetoric* for the details of his psychology. He merely poured a traditional content into a mechanical mold that he adapted from Galileo, Pierre Gassendi, and the ancient atomists. The Schoolmen provided the thinkers of the seventeenth century with something solid and disciplined to revolt against. And, as with most rebels, these thinkers were really revolting against a mass of assumptions that were deeply embedded in their own consciousness. Indeed, in a certain sense their revolt was only a return to other elements in their intellectual heritage—the precipitates left by the Pythagoreans, Plato, and the atomists.

**DESCARTES.** Descartes's view of the mind was a return to Plato, enriched by the introspective musings of Augustine and made more precise by developments in the natural sciences.

**Nature and mind.** The natural sciences had made leaps forward not because of a vast accumulation of new facts, though one of the features of the Renaissance had been man's turning his gaze out toward the natural world; it was, rather, because of the amazing success that had attended the application of geometry to the phenomena of the natural world.

The success of geometric thinking about nature tended to corroborate what Plato had said about the status of reason as contrasted with the senses; it also con-

vinced the new natural philosophers like Johannes Kepler, Galileo, and Descartes that the real qualities of the natural world were those which could be treated geometrically. Matter was homogeneous, as the atomists had said. Qualitative distinctions, which had been exalted by Aristotle into irreducible natural kinds, were appearances of the varying motions and configurations of the underlying bodies. The Aristotelian doctrine of form and matter was banished; so were the final causes that he had postulated in nature.

How, then, was mind to be conceived, once the Aristotelian doctrine of form and matter had been discredited? There were two obvious possibilities. One was to adopt Epicurus's view that soul and mind were configurations of light and mobile atoms. The other was to revert to the Platonic view that mind is an altogether different type of substance that inhabits the body. Descartes adopted the second course, partly because he shared Plato's view about the wonder of reason and its difference from sensation and bodily processes and partly, no doubt, because of his Christian convictions about God, freedom, and immortality.

**Mind.** Descartes's departure from Aristotle was much more radical in his account of the soul than in his account of the mind. Whereas Aristotle had described the soul, even in its most primitive manifestations, in teleological terms, Descartes attempted to describe all its lower functions, which were connected with the body, mechanically. His account of mind was not dissimilar in its main outlines from Aristotle's account of reason, which was the most Platonic part of his doctrine, for both accounts held that mind comes from without, furnishes the ultimate principles of thought, and may be considered apart from the body. Indeed, Descartes stated emphatically that the mind can think without a body.

For his account of mind Descartes looked into himself in the manner of Augustine, but he rejected that reliance on faith which was epitomized by the protestation *Credo quia absurdum* ("I believe because it is absurd"). Nothing that was not clearly and distinctly present to the mind was to be included in a judgment. Everything must be doubted—even mathematical truths—until a belief can be found that applies to what exists and that it would be self-contradictory to deny. Descartes's *cogito ergo sum*—his more precise rendering of Augustine's intuitive certainty about his existence as a thinking being—was the result.

Descartes explored the rest of what was intimated in this first certainty and tried to spin out of it all sorts of other truths—for example, the existence of God and of

an external world. The details of his attempted demonstration do not concern us here. They effectively established, in Descartes's view, the existence of thinking substances that were innately so constituted that they would come to form clear and distinct ideas of extension, figure, motion, and other simple natures. Ideas are all mental; as images they are presented through bodily processes, images being apparently corporeal.

Minds were thought to be passive in cognition. When a mind is thinking clearly and distinctly, its ideas correspond to the real qualities of objects. But minds are also active in volition. At the intellectual level their activity consists only in assent to the necessary connection between ideas, and volition is one of the most potent sources of error, for there is often assent when ideas are not clear and distinct. Volition is also the cause of action and is operative in attention, recollection, and fantasy.

**Body-mind relation.** Descartes's account of the body-mind relation was not dictated solely by Platonized Christian piety. It was equally the product of his knowledge of science and his convictions about scientific method. First, Descartes was convinced that the body is a machine and that animals' behavior could be explained mechanically, animals having no souls. He was acquainted with the discoveries of William Harvey that showed the circulation of the blood to be a mechanical process. Furthermore, mechanical models were a feature of the age. Decorative fountains were constructed with model men that were moved hydraulically and even uttered sounds like words. Descartes thought that the body contained tubes like water pipes along which the animal spirits (the up-to-date rendering of Aristotle's "connatural spirits") coursed. Because many movements of the body can be executed without conscious intentions, Descartes assumed that these could be explained in the same way as the movements of the hydraulic men. He has thus been credited with the discovery of reflex actions. He thought that all animal behavior could be explained in this way.

Second, Descartes believed in the principle of conservation of energy. The quantity of motion imparted to and conserved in a system being constant, there could be no extra source of energy deriving from volition. Thus, the relationship between body and mind had to be conceived in a way that was consistent with this principle.

Third, Descartes held that scientific explanation consisted of making deductions from relations grasped between clear and distinct ideas. Clear and distinct ideas were available of the simple natures of body (for example, extension, figure, motion) and of mind (thinking, will-

ing) but not of the relation between them. Descartes held fast to the obvious fact that body and mind interact (for when I will, it is my arm that moves; I feel pain when my body falls and not when a stone falls). But we have only a confused idea of this interaction. His account of the relationship between them was therefore only a likely story with which he was not really satisfied. It only narrowed down the point at which the crucial philosophical difficulties occurred.

Descartes knew that muscles operate in opposing pairs and that nerves are necessary for sensation and movement. He pictured nerves as tubes along which animal spirits flow. Changes in the motion of these animal spirits cause them to open some pores in the brain rather than others. When this happens, the spirits are deflected into muscles that move the body by being distended laterally and, thus, shortened. At the level of instinct and habit this process is purely mechanical. At the level of conscious intention, however, something more had to be postulated, the impact of mind on body at the crucial switching point of the spirits, the pineal gland.

Descartes supposed that in sensation motion was transmitted from the stimulus object through a medium to the sense organ and thence along the spirits in the nerves to the pineal gland in the center of the brain, where an impression was made like that of a seal on wax. This was a material image that stimulated the soul to produce a corresponding idea. Descartes gave a similar account of passions in the narrow sense of emotions and organically initiated disturbances, which have their source in the agitation of the spirits. By passions in a general sense, Descartes meant all things that *happen* to minds, including sensations, lower forms of memory, feelings, emotions, and other disturbances of reason. These he contrasted with the mind's activity. All such incoming stimuli generally give rise to an act of will. Willing again makes contact with the body at the pineal gland, and a chain of events is started in the body terminating with the movement of the muscles, which produces voluntary action.

The soul is like a pilot in a ship in that it can effect the direction but not the amount of bodily movement. Thus, Aristotle's image of active reason could be reconciled with the principle of the conservation of energy. Descartes's hypothesis that interaction between body and mind occurred at the pineal gland did nothing to dispel the philosophical perplexity about how this interaction could be conceived, and then the pineal gland later was shown to be nothing more than an obsolescent eye. Descartes was attached to this idea because the pineal



gland was the only part of the brain that was not duplicated in both halves of the brain. He was convinced that the soul, being unitary, could not affect the body at two points. His hypothesis enabled him to keep his mechanistic account of the body intact.

For a long time it has been fashionable to deride Descartes's rather disastrous form of dualism and even to suggest that he created the body-mind problem. This is a piece of intellectual insularity. Descartes was perhaps the first thinker to formulate the problem at all clearly. It would be possible to deny his basic assumption that body and mind are qualitatively distinct substances and still to claim that apart from this metaphysical extravagance his statement of the problem brought out at least two cardinal points that are involved in it. First, he obviously saw the logical incongruity of explaining mental processes, such as geometric reasoning and deliberating before action, in mechanical terms. There is a logical gap between the types of explanation used, as Aristotle had pointed out in his criticisms of the mechanists who held that the soul was moved. Descartes, in his account of the transactions that were alleged to take place at the pineal gland, must have thought that motion at this point is somehow identical or correlated with the mental activity involved in producing an idea or making an act of will. His hypothesis did much to draw attention to this logical disparity between the two types of description.

Second, Descartes's account did much to establish privacy, rather than Aristotle's criterion of purpose with plans and rules superimposed at the level of the rational soul, as the main hallmark of the mental. As has been indicated, Descartes's theory in this respect marked the culmination of a trend that can be traced back through Augustine and Plotinus to Philo. To attribute mind to something is not just to say that men act in accordance with rules and that their movements persist toward ends. It is to say that they act like this because of their knowledge of rules and because they are conscious of ends. Consciousness is crucial for picking out the obvious respect in which men differ from cunningly contrived machines. Descartes must be credited with the clearheadedness to have stood firm on this cardinal point.

**SPINOZA.** Benedict de Spinoza's system was a consequence of pushing Descartes's assumptions to their logical conclusions.

*Nature and mind.* Descartes had accepted the traditional notion of substance as that which is a cause of itself, can be conceived through itself, and needs only itself in order to exist. Spinoza (1632–1677) argued that if

this is the definition of substance and if there is such a substance, there can be only one such substance, which can be called either nature or God. Nature, so conceived, must have infinite attributes, but we know only two of them, thought and extension. God is therefore "the place of the world and the whole system of thinking." Everything is a mode or modification of God. Thus, nothing can be adequately explained unless its occurrence can be deduced from principles applying to the system as a whole.

Explanation is deductive in character and accords with mechanical principles. Unlike Descartes, Spinoza envisaged a science of psychology in which mental as well as physical phenomena could be deduced from quantitatively expressed laws. Emotions, he argued, must obey laws just as lines, planes, and bodies do. Human beings, as part of nature, must exhibit the general characteristics of all modifications of God or nature. They must be determined within a system; they must have a mental and a physical aspect; and they must exhibit *conatus*, or the striving to persist within their own being. These characteristics must now be considered in turn.

In stating that human behavior was determined within a system, Spinoza wished to oppose what he considered to be two basic illusions that human beings had with respect to themselves. The first of these was the illusion of free will. People are convinced that they have free will, he argued, because they are conscious of their actions but ignorant of their causes; thus, they conclude that they are uncaused. If stones were conscious, they, too, would believe in free will. Yet human behavior can be explained just as can the movements of stones. In both cases the explanation will consist in deducing what occurs from the laws of the system of which they both are part, ultimately the system of nature as a whole. The human body is a system of simpler elements maintained in an equilibrium, but this system is part of a broader system, not a self-contained isolable system. Adequate explanation is seeing events as part of the whole system of nature; in this system there are no final causes. Nature just is, like a vast, timeless machine.

*Body and mind.* How then was the body-mind relation to be conceived? Spinoza was one of the first to point to the difficulties in Descartes's pineal gland hypothesis. Spinoza's solution was to suggest that interaction does not take place for the very good reason that body and mind are correlated attributes of the same underlying substance, not distinct substances. Indeed, Spinoza says that the mind is the idea of the body. This is obvious enough at the level of immediate confused ideas that are

of bodily states. But the changes in a man's body are part of a larger system, which includes the properties of the food absorbed in nutrition. A wider knowledge of the events in a man's stomach is possible for a physiologist who can understand the laws governing them. He would see these events as part of an ever widening network of events which constitute nature. The man's feeling of stomachache, on the other hand, would be confused, fragmentary, and inadequate, an idea of an effect cut loose from its causes.

This illustrates the difference between what Spinoza called the first and second grades of knowledge. The materials of the first grade are the confused ideas of bodily states that we call feelings and sensations. These ideas are connected only by principles of association. This is the level of sense perception and imagery, of uncritical beliefs founded on animal instinct, association, and hearsay. The second grade of knowledge is rational insight. At this level rational connections are grasped as general notions develop that connect an ever widening system of events. The more abstract and general thought becomes, the nearer it approaches the thought of the Cartesian physicist and, ultimately, God's thought. There is also a third grade of knowledge, called *scientia intuitiva* by Spinoza, which is more mystical. It is a return from the abstract laws of the scientist to a grasp of the particular as illuminated by such laws. The role of the body, as that which is correlated with mind and of which mind is an idea, seemed to recede when Spinoza passed to reason, or the second grade of knowledge. Mind as the idea of the body becomes at this point almost as difficult a notion as Descartes's notion of mental activity somehow mirroring movement in the brain, for thinking is not of or about body or brain states any more than it is a form of movement which is similar to or identical with brain states.

*Conative aspect of mind.* Spinoza's account of mental phenomena was much less intellectualistic than that of Descartes. Indeed, in certain respects he reverted to Aristotle's emphasis on teleology and self-maintenance. Spinoza held that the most important characteristic of every modification of nature was its conatus, its striving to persist in its own essence. In man, as in every other natural modification, there is an inherent tendency to react to all changes in a way that maintains its characteristic unity and equilibrium. A person differs from animals in being self-conscious in this endeavor.

Spinoza employed this homeostatic postulate to rewrite Descartes's account of the passions as presented in *Les passions de l'âme*. Descartes had paid particular attention to the causal influence of animal spirits and had

left rather vague the part played by the cognitive grasp of the situation, though he generally put forward an ideomotor theory. Spinoza evinced little interest in the physiology of the matter. Instead, he developed a theory of motivation by harnessing Descartes's passions to his own homeostatic principle. He postulated that whenever a body is acted on by another body, its vitality may be increased, may be diminished, or may remain constant. The awareness of these occurrences is the mental aspect of the psychophysical states which are called emotions. There are thus three primary emotions corresponding to increase, diminution, or maintenance of bodily vitality. These are joy (*laetitia*), grief (*tristitia*), and desire (*cupiditas*). As a result of experience people tend to keep before them what will increase their vitality and remove what will decrease it. "Love" is thus defined as "joy accompanied by the idea of an external cause."

Spinoza drew a sharp distinction between the passive emotions which characterize the first grade of knowledge and the active ones which mark the second and third grades. People are passive when the cause of changes in them lies outside them. In this state of human bondage the emotions that accompany confused, fragmentary ideas are thrust on people; they tend to be sporadic, inordinate, unpredictable, and obsessive. Individuals are subject to panic, jealousy, and overmastering loves and hates. When a man passes to the second grade of knowledge, however, his vitality is increased, and there is a distinctive form of joy that goes with the use of reason. The explanation of human conduct is now to be sought within him, in his clear understanding of the world and of his relation to it. By understanding himself, including his own emotions and history, as part of the system of nature, a man can attain a kind of freedom, which depends upon his acceptance of his own nature. He is then capable of rational self-love and rational benevolence and can attain glimmerings of the greatest good which he can possess—"the knowledge of the union which the mind has with the rest of nature." The attainment of this state brings its own delight.

In making suggestions for attaining this state of blessedness, Spinoza in many respects anticipated later psychoanalytic techniques, as well as the general psychoanalytic aim of replacing subservience to irrational promptings by rational control based on self-knowledge. He thought, for instance, that many irrational reactions could be traced back to an early reaction to an object to which the present object had become associated by irrelevant similarities. Scientific understanding of this might help to dissociate the emotion from the irrelevant stimu-

lus. He was not so naive, however, as to suppose that mere intellectual understanding could free an individual from the obsessiveness of emotion. It takes an emotion to master an emotion. And Spinoza thought that seeing things “under the aspect of eternity” had a specific emotional accompaniment. Hence, the psychological shrewdness as well as the ethical profundity of his remark, “Blessedness is not the reward of right living; it is the right living itself. Nor do we delight in blessedness because we restrain our desires. On the contrary it is because we delight in it that we restrain them.”

**HOBBS.** Hobbes (1588–1679) already subscribed to the deductive model of geometry when he visited Galileo in 1636. He returned replete with concepts and laws that were to form the foundation of his psychology. For the idea had dawned on him, perhaps suggested by Galileo, of applying the new natural philosophy to human behavior. Of course, Epicurus had long ago sketched a mechanical theory of mind, but it was very general. Galileo had worked out the details of a new theory of motion. Could not still further consequences be deduced from the law of inertia? Harvey had deduced the theory of the circulation of the blood from mechanical postulates. Could not Hobbes apply the details of this new theory of motion to psychology and politics?

*Body and mind.* Hobbes did not really see any particular problem about the relationship between body and mind because for him everything was body. Even God must have a body if he exists, for “substance incorporeal” is a contradiction in terms.

Thus, “conceptions and apparitions are nothing really but motions in some internal substance of the head.” Sensation is “some internal motion in the sentient,” and pleasure is “nothing really but motion about the heart.”

In truth, Hobbes was not much worried by such philosophical niceties as whether, according to his theory, mental phenomena like thinking were being postulated as identical with or merely causally dependent on motions in the head. He was much more interested in working out a mechanical explanation of these phenomena. This is what makes his psychology of absorbing interest. It represents just about the first attempt in the history of psychology to put forward in any detail something that begins to look like a scientific theory.

*Mechanical theory of mind.* According to Hobbes, in sensation the sense organs were agitated by external motions without which there could be no discrimination and, hence, no sensation. The selectivity of perception

was explained by suggesting that while a sense organ retains motion from one object it cannot react to another; similarly, in attention the motion from the root of the nerves persists “contumaciously” and makes the sense organ impervious to the registering of other motions. Imagination was explained by a strict deduction from the law of inertia: “When a body is once in motion, it moveth, unless something else hinder it, eternally; ... so also it happeneth in that motion, which is made in the internal parts of man, then, when he sees, dreams, etc.... Imagination therefore is nothing but decaying sense.” This decay is not a decay in motion, which would be contrary to the law of inertia. It comes about because the sense organs are moved by other objects. This explains why dreams are so vivid, for in sleep there are no competing motions from the outside world. Thus, the longer the time that elapses after sensing an object, the weaker the imagination. Memory is imagination with a sense of pastness added to it.

This was an exciting and an ingenious theory. The difficulty about it is that the type of distinction implied in the explicanda cannot really be deduced from the mechanical postulates of the theory, for the differences between perceiving, imagining, and remembering are basically epistemological ones implying standards and criteria different from those that might be attributed to mere movements. Hobbes never faced the basic difficulties that Aristotle first formulated in his opposition to the theory that the soul was itself moved. Nevertheless, Hobbes did produce something that looked like a scientific theory. Its conceptual difficulties attend all psychological theories that attempt to translate epistemological distinctions into differences of process.

*Mechanical theory of action.* In the theory of action Hobbes attempted to get rid of final causes and to substitute efficient causes for them. To do this, he had to introduce the concept of endeavor, which was very different from Spinoza’s conatus. He used the term *endeavor* to designate infinitely small motions, which he postulated as occurring in the medium between the object and the sense organ, between the sense organ and the brain, and heart. His theory of motivation was that external objects transmit motions by a medium to the sense organs and from there to the brain and to the heart; this results not only in the production of images but also in some alteration or diversion of vital motions round the heart. When these incoming motions help the circulation of vital motions, it appears to us as pleasure, and the body is guided to preserve the motions by staying in the presence of the stimulating object; and conversely with pain.

Appetite and aversion are thus the first endeavors of animal motion. They are succeeded by the flow of animal spirits into some receptacle near the “original” of the nerves which brings about a swelling and relaxation of the muscles causing contraction and extension of the limbs, which is animal motion.

Hobbes thought this mechanical account of action was quite consistent with ascribing a central role to consciousness, for in Hobbes’s view all action was voluntary in the very strong sense that it is preceded by the thought of an end to be attained. He also claimed that the only way to develop a science of human nature was to look into ourselves and analyze what we find there. Hobbes found two basic motions of the mind, “the one arising from the concupiscible part, which desires to appropriate to itself the use of those things in which all others have a joint interest; the other proceeding from the rational that teaches every man to fly a contra-natural dissolution, as the greatest mischief that can arrive to nature.” Everything we do is derived from the desire for power or the fear of death. Conflict between manifestations of these basic motions of the mind leads to deliberation. In this “alternate succession of appetite and fear” the one that emerges triumphant is called “will.” “Will therefore is the last appetite in deliberation.” Free will is an illusion, for the outcome of such conflicts can be explained mechanically.

*Theory of passions.* On top of this mechanical ground plan Hobbes superimposed an account of the passions taken largely from Aristotle’s *Rhetoric*. They are to be distinguished by reference to the objects of appetite and aversion as well as by our opinion of attaining such objects. Ambition, for instance, is desire for office; hope is appetite with an opinion of attaining. Individual differences are due, in the main, to differences in the mobility and agility of the animal spirits. Dullness, for instance, derives from “a grossness and difficulty of the motion of the spirits about the heart.” Hobbes even had a theory of laughter, which he thought to be the expression of sudden glory caused by something new and unexpected in which we somehow discover ourselves superior to others.

Hobbes assigned a special place in his theory of the passions to curiosity, which, together with the ability to name things and hence to reason deductively, distinguishes humans from animals.

Hobbes’s account of the passions was unusual in that it was so positive. For him passions were not, as for the Stoics, imperfect reasonings; they were a particular case of motion in the natural world on which his account of human nature was erected. Nevertheless, when he dealt

with what was distinctive of man, his reason, Hobbes parted company with both naturalism and mechanical theory. The type of reason, called prudence, which enables man to satisfy his desires more efficiently, on the basis of experience, must be sharply distinguished from the reason by means of which men are able to arrive at the universal truths of geometry and philosophy.

*Scope of mechanical theory.* This is not the place to enter into the tortuous details of Hobbes’s nominalist theory of meaning or his conventionalist theory of truth. It is important to note, however, that in dealing with these specifically human facets of behavior, just as in his treatment of the foundations of civil society, Hobbes defended a position that stressed above all the role of artifice and convention. He even put forward a kind of contract theory of definition to parallel his social contract theory of government. These accounts were underpinned by a very crude causal theory of signs as well as by a mechanical theory of human nature. But no clear connection was ever made between the conventionalist and naturalistic elements. David Hume later tried to make such a connection by suggesting that reason was a wonderful and unintelligible instinct in human nature. Hobbes, however, more or less ignored his own mechanical theory when he dealt with geometry, law, logic, and other such artificial creations of human reason.

Thus, although Hobbes was the first thinker to develop in any detail a mechanical theory of mind, he also, more or less unwittingly, exhibited the glaring difficulties in such an undertaking. Indeed, the things in which he was most interested, apart from politics, were precisely those things which it is very difficult to accommodate within a mechanical theory.

LEIBNIZ. Gottfried Wilhelm Leibniz (1646–1716) understood much better than Hobbes the new natural philosophy; indeed, his discovery of the infinitesimal calculus contributed considerably to it. However, he resisted its mechanistic implications. Descartes had viewed nature, the animal world, and bodies as machines but had stopped short at mind; Hobbes had mechanized mind as well. Leibniz went to the other extreme and mentalized nature. In many respects he reverted to Aristotle.

*Nature and mind.* The *Monadology* was a brilliant synthesis of Aristotelian logic taken seriously and a variety of trends in the natural sciences. The whole Cartesian philosophy presupposed the subject-predicate view of judgment in which every proposition, when reduced to logical form, has a subject and a predicate. Moreover, the predicate was thought to be contained in the subject. The

Aristotelians thought that this common structure of language mirrored a world of substances composed of various attributes. Leibniz, like Spinoza, took the definition of substance seriously; he thought that it was the cause of itself, could be conceived by itself, and needed only itself in order to exist. But where Spinoza concluded that if this was the definition of substance, there could be only one—namely, God or nature—Leibniz concluded that the world must be composed of countless substances all exhibiting the features picked out in their definition. These monads develop according to an immanent principle that is their force or essence. Everything that will ever happen to them, their predicates, is included in their original notion. The principle of sufficient reason explains the succession of these states in time, the identity of a substance at different times being recognized by “the persistence of the same law of the series.” Now I am a substance and know by introspection that I am characterized by appetite and perception. What I know about myself must in general be a paradigm for the basic structure of all substances. But no two substances are alike. In perception they all mirror the universe from a particular point of view. There is no interaction, however. Each monad is windowless and develops because of its own immanent principle, not because of external causal influences. The monads seem to influence one another only because of the preestablished harmony of their immanent development.

This bizarre application of an ancient logical doctrine to the world accorded nicely with various new developments in the sciences. Leibniz naturally regarded it as consistent with his discovery of the infinitesimal calculus, the guiding idea of which was that a succession of states develops according to a law governing the series. The successive states of a monad flow into one another like a series of terms differing infinitesimally, their development being defined by the law of the series. This fitted well with the law of continuity, which held that *natura non facit saltus* (“nature makes no leaps”). Change is a summation of infinitesimal degrees of change. Furthermore, the recent discovery of the microscope revealed that if a piece of cheese or a seemingly empty pool is examined, each will be found to be teeming with life. Could not all nature, therefore, be alive—a vast system of monads at varying levels of development? In embryology, too, the doctrine of preformation was in vogue. The assumption that all the characteristics of an adult animal exist in embryonic form from the moment of generation supported Leibniz’s view that from the original notion of the monad all its later states and characteristics could be deduced. His conception of the essence of monads being

force or activity was connected, too, with his contribution to the dispute in dynamics about the relationship between force and mass. Leibniz held that his concept of *vis viva* or activity directed toward the future states of the monad was required by his discovery of the conservation of momentum.

The synthesis of Aristotelian logic and these trends in science made Leibniz utterly opposed to the mechanistic picture of nature and of man in which the real world was a world of bodies in motion having only primary qualities whose changes were to be explained only by reference to efficient causes. What is real, he claimed, is not what is mathematically measurable but our experience of activity and perceiving. Nature, as well as man, is characterized by appetite and perception. Final causes are reconciled with the laws of motion by the principle of sufficient reason, which governs the unfolding of the immanent nature of the monads. The difference between substances is only one of degree of clarity in perception and of self-consciousness in appetite. Bare monads have a minimum of perception and appetite. Their perception is confused, and their appetite is blind. Souls, or conscious monads, have memory, feeling, and attention. Animals, or, rather, the dominant monads of animals, are examples. Rational souls, or spirits, are self-conscious; unlike brutes, which are “empirics” and are aware only of particulars, they can reason and understand necessary truths. Extension is only an appearance, the way in which low-grade monads appear to us; the laws of motion are just appearances of the laws of appetite which depend ultimately on God’s choice of what is best. Aristotle and Galileo are reconciled, but Galileo’s and Isaac Newton’s laws are, at best, laws of appearances.

*Concept of mind.* Leibniz’s concept of mind or soul was articulated in what he said about perception and appetite. He regarded perception as marvelous because it cannot be conceived of as an action of the object on the percipient, for the monads are windowless. Perception is better regarded as the expression of a plurality in a unity. One thing may be said to express another when there is a constant and regular relation between what can be said about the one and about the other. It is thus that a projection in perspective expresses its original. The monads are perspectives of the universe from different points of view. Expression is thus the genus of which perception, animal feeling, and intellectual knowledge are species.

Leibniz combined this highly metaphysical account of perception with some shrewd objections to John Locke’s tabula rasa theory of the mind. He held that the senses provide us only with instances and by themselves

cannot provide the sort of universal knowledge that we have in science. The mind is active and categorizes experience by means of which it interprets the testimony of the senses. The proper analogy for the mind is not a tabula rasa but a block of veined marble. In this doctrine Leibniz harked back to Aristotle's active reason and laid the foundation for Immanuel Kant's categories. Locke, he argued, had in fact tacitly admitted this in postulating mental operations that are known by reflection.

Leibniz maintained that Locke was wrong in saying that the mind does not always think. We have an infinite number of perceptions of which we are not aware. Habituation and wandering attention, as well as the smallness of the perceptions, explain our failure to notice them. Our attention is often drawn to a sound that has just occurred and that we would not otherwise have consciously noticed, although we registered it. "These insensible perceptions are also the signs of personal identity and its constituents; the individual is characterized by traces of his previous states which these perceptions preserve by connecting them with his present state." They are also the means of recollection. They explain decisions that seem arbitrary to us, like turning to the left rather than to the right; they explain frequent feelings of uneasiness which are not intense enough to be felt as pain. These insensible perceptions, he argued, are "as much use in pneumatics as is the insensible corpuscle in physics." Both are beyond the reach of our senses, and there are as good grounds for believing in one as in the other. Since "nature makes no leaps," these insensible perceptions must accord with the law of continuity. "All this brings us to the conclusion that observable perceptions come by degrees from those which are too small to be observed."

Although Leibniz confused some rather different things in this doctrine—for example, unconscious perceptions, minute perceptions that summate like the noise of waves in the roar of the sea, and confused perceptions—he prepared the ground for the concept of unconscious mental processes which was to prove so important in nineteenth-century thought, and he anticipated later investigations of subliminal perception and "determining tendencies." This shows how a highly speculative theory can lead to the emphasis on facets of experience which may be very important but which have previously been disregarded.

Leibniz's emphasis on appetite as the other main characteristic of monads was a welcome change from the intellectualism of Descartes and Locke. However, Leibniz made no detailed empirical derivations from this notion to match the derivations made from his concept of per-

ception. It had more affinities with Spinoza's "conatus" than with Hobbes's "endeavor," although it was really the Aristotelian conception of the formal and final cause brought up to date and made compatible with dynamic theory. His concept can best be elucidated by quoting him; he calls his concept by the Aristotelian term "entelechy," which is "a power mediating between the simple faculty of acting and the definite or effected act. It contains and includes effort. It is self-determined to action, not requiring to be aided, but only requiring not to be inhibited. The illustration of a weight which stretches the cord it is attached to, or of a bent bow, may elucidate the notion."

*Soul and body.* Leibniz believed that every living creature is composed of a vast number of special organic structures each developing in its own characteristic way; they are all so coordinated and mutually complementary, however, that together they act as an individual. The unity is the soul or the dominant monad; the multiplicity is the body or assemblage of bare monads. The monads of the body all have their own activity, and they are represented or mirrored in the perceptions of the dominant monad or mind. The mind has no power to interfere with or penetrate the forces that it seems to direct. The activities of the monads of the body subserve the dominant activity of the mind as the players of an orchestra, each playing independent parts, subserve the performance of the symphony, and the symphony is the resultant harmony, which has been preestablished. The manifold activities of the bare monads thus combine to bring about the end of the dominant monad. The body depends on the mind in the sense that the reason of what happens in the body is to be found in the mind (compare to Aristotle's view of soul and body).

Thus, Leibniz reverted to a view of mind and nature which was basically Aristotelian, but he transformed the Aristotelian entelechy by giving it the basic hallmarks of Cartesian mind—thinking and willing as experienced from within. Furthermore, he pressed the emphasis on privacy much further than Descartes by claiming that the monads are windowless and that everything that will ever happen to them is contained in their original notion.

There was, however, another radically different concept of mind which developed out of Descartes's stress on privacy and incorrigibility as the hallmarks of mental states. This was that of British empiricism, which culminated in Hume and the associationists.

**HUME.** The contribution of Hume (1711–1776) to psychology was not very extensive in its details because his

theorizing about the mind, like that of George Berkeley and Locke, was mainly a way of doing epistemology. And there were special reasons, deriving from his epistemological position, for his eschewing speculation about the relationship between mind and body and the general status of mind in nature. Nevertheless, his general concept of mind was of considerable historical importance. It was the first thoroughgoing attempt to eliminate spiritual substance altogether, and it was the first theory to make reason subservient to the passions and to extol the importance of instinct and habit. It was also the first attempt to develop a Newtonian theory of mind and to erect the principles of the association of ideas into scientific postulates—an undertaking which considerably influenced David Hartley and hence the course of associationist psychology.

*Hume's predecessors.* John Locke (1632–1704) took from Descartes the assumption that we are confronted with our own ideas, not with things, and that some kind of certainty is both desirable and attainable. He rejected, however, Descartes's doctrine of innate ideas and adopted a Baconian version of empiricism. He postulated simple ideas of sense that made their imprint on the passive tabula rasa of the mind. Once ideas got into the mind, Locke's theory more or less followed Descartes's, for he believed that the active spiritual substance within intuitively relations between ideas, the relations which form the foundations of knowledge. Locke, however, did not stick consistently to his "way of ideas." For example, he asserted, like Descartes, that we have intuitive knowledge about our own existence as selves and "sensitive" knowledge of things existing independently of our perceptions of them. They are material substances that support "powers" to produce in us ideas of primary qualities, which are real properties of the things in question, and secondary qualities which are not real.

George Berkeley (1685–1753) stuck more consistently to the way of ideas and eliminated material substance, of which we have and could have no idea because it is a logical absurdity; the representative theory of perception; and the distinction between primary and secondary qualities. He claimed, however, that we have "notions," rather than ideas, of ourselves as active agents and of other minds, including God. We also have a notion of our own causal activity. Berkeley relied on this notion to distinguish ideas of sense from ideas of imagination, for having eliminated the concept of a thing independent of our perceptions, Berkeley had to have a criterion for distinguishing what are commonly called things from the mere coexistence of qualities; imaginary objects, for

instance, appear to us as clusters of coexisting qualities. Thus, he claimed that when we see objects, it is God talking the divine sense language and producing ideas in our minds; when we imagine objects, we are doing the producing ourselves and have a notion of our own agency in so doing. Berkeley's stress on the activity of the mind contrasted strongly with Locke's tabula rasa.

Hume simply stuck rigorously to the way of ideas and eliminated Berkeley's "notions." There was no simple idea of material substance, of ourselves and others as spiritual substances, of God, or of causal agency. All that was left, therefore, as genuine components of the mind were ideas themselves and certain links between them. Hume likened the mind to a theater "where several perceptions successively make their appearances, pass, repress, glide away," and to a political organization in which the members come and go but the principles of organization—the principles of the association of ideas—persist.

*Hume's contributions.* Hume was the first to attempt an explicit distinction between images, which he called impressions, and what we would now call sensations—he called them ideas. He regarded them as two sorts of perceptions. Impressions could not be distinguished from ideas in a Lockian way by their relation to an external object. For Hume, following the way of ideas, disclaimed any possibility of knowledge of a world of objects existing independently of our perceptions. And, because he ruled out notions, Berkeley's appeal to awareness of our causal agency in producing ideas of imagination was not open to him. Of course, like Berkeley, Hume agreed that what we call things exhibit a certain constancy and coherence; they resemble past clusters of qualities. We assume independent existence in order to connect past with present perceptions. But, he argued, we can no more demonstrate the existence of a world independent of us than we can demonstrate that pleasure is preferable to pain.

There are, however, subjective criteria for making the distinction between images and sensations, which is all that remains once belief in a world of independent objects has been ruled out. These are the criteria of vividness and order. Hume suggested that ideas could be picked out because they were faint copies of previous impressions. In other words, impressions are both more vivid than ideas and prior to them. But he gave counterexamples to both these criteria—those of vivid ideas in fever or madness and of forming an idea of a color that had never previously been presented as an impression. In the case of fever or madness Hume suggested that the imagination transfers the vividness of an impression to

an idea. Similarly, our belief in an external world is a work of the imagination.

Hume's recourse to the imagination was of cardinal importance in his account of the mind because it linked his theory of knowledge with his rehabilitation of feeling. It has often been remarked that one of the main features of Hume's philosophy was a reversal of the roles hitherto ascribed to reason and feeling. He brought over into epistemology his ethical theory, which he adapted from Francis Hutcheson's theory of moral sense, that moral judgments are based on feeling. "Reason is, and ought always to be, the slave of the passions." This moral sense was the product of biological properties inherent in the species; it had its counterpart in our judgments of matters of fact and existence. Reasoning is "nothing but a wonderful and unintelligible instinct in our souls." Our belief in the reality of causal connections or in the existence of an external world or that the future will resemble the past are instinctive and indemonstrable. "Nature, by an absolute and uncontrollable necessity, has determined us to judge as well as to breathe and feel." The categories used by scientists in their theories, such as continuity and causality, are largely products of the imagination.

Hume stressed facets of human nature that had been largely neglected since Aristotle. He postulated an original fabric of human nature consisting of various propensities not unlike that of later instinct theorists. He also extolled the place of habit in conduct, not simply in explaining such developed forms of behavior as obedience to government but also in explaining the origin of some indemonstrable beliefs. For instance, he held that the idea of causal connection could be analyzed into the elements of priority in time of event *A* to event *B* and constant conjunction of event *A* with event *B*, together with a conviction of the necessity that *B* must follow *A*. As there was no impression of this necessity given in experience, Hume attributed our belief in it to habit or a "determination of the mind" brought about by experience of such constant conjunction and the force of the imagination.

*The passions.* Appropriately enough, the details of Hume's psychology consisted mainly of an elaborate and highly complex theory of the passions, stated in Book 2 of his *Treatise of Human Nature*. One of Hume's tasks was to rehabilitate the passions, the natural feelings of decent people, from the Puritans' distrust and the rationalists' disregard. He also had to demolish sophisticated theories, deriving from Hobbes, in which all passions were regarded as forms of self-love. Whereas Bishop Butler attacked psychological hedonism in order to establish the

supremacy of conscience, Hume refuted the hypothesis of self-love in order to make way for his rival hypothesis of innate benevolence and sympathy.

He also regarded the sensations of pleasure and pain as part of the original fabric. In a passion one of these sensations is accompanied by an affection. The direct affections include desire and aversion, joy and grief, hope and fear. The difference between these depends on the character of the expectation of good or evil. Desire is for present good, joy for assured good in the future, and hope for probable though remote good in the future. Hume thought that through experience these affections, together with the sensation of pleasure or pain associated with them, can become associated with an object. This generates such indirect passions as pride and humility, when the object is ourselves, or love and hate, when the object is other people. Benevolence and malevolence, however, are not derived from love and hate. Hume classed them as direct and instinctive.

Sympathy occupied a role in Hume's theory of passions somewhat similar to imagination in his theory of belief. The idea of another person's feeling is said to be associated with the idea of oneself, and the required liveliness is thus imparted to the otherwise neutral conception of another person's joy or sorrow.

The idea of the self played an important part in Hume's intricate account of the passions. Like the idea of causality, it presented a serious problem for analysis, for we believe strongly in the reality of both of them. Yet, Hume argued, there was no simple impression of sense from which these ideas derived. Introspection revealed only "some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure." What we call self must therefore be "a bundle of perceptions." Like Locke, Hume then went on to compare the self to an oak, a vegetable, or any type of organism which maintains itself through change by virtue of its relations. Another apt analogy is the self-maintained unity of a political association. But Hume maintained that the unity of this bundle, which makes it a "connected heap," is associative, not real; there are no grounds for ascribing to it the simplicity and permanence which are required for real unity. Perceptions are loose, separate, perishing existences. There can be no real links between them. The problem is to explain how we come to believe that there are.

Hume made the same type of move in relation to the idea of self that he made in the case of causality. He demonstrated that if the way of ideas is followed, there is no ground in experience for believing in the reality of the self; he then embarked upon some speculative psychology



to explain how we come to have this belief. He suggested that members of the bundle are related to one another in a specific way in time, the order being preserved by memory. The members have the relations of resemblance and cause and effect between them. But cause and effect is not a real relation; thus, no real unity characterizes the self. We come to believe in it because of the “felt smoothness” with which we pass from one idea to another once the associative links have been established.

*Nature and mind.* Although Hume’s adherence to the way of ideas ruled out wide speculations about the place of mind in nature, there was a highly imaginative idea behind his positivistic system. Hume regarded himself as the Newton of the sciences of humankind. He made frequent references to his pursuit of the experimental method and thought his rigorous interpretation of the way of ideas to be thoroughly consistent with Newton’s methodological canons of economy and simplicity in explanation, testability of hypotheses, and refusal to postulate occult causes. Hume stressed that once we have arrived at the original fabric of human nature, it is futile to attempt to satisfy any further our intemperate desire to search for other causes.

But Hume did not emulate Newton merely in his methodology. He also regarded his concepts in the psychological sphere as parallel to Newton’s concepts in the physical. His simple impressions were the equivalent of Newtonian atoms, and his principles of association were likened to the “gentle force” of Newton’s principles of gravitational attraction. Indeed, Hume regarded imagination and, perhaps, sympathy as cohesive forces. When imagination works according to the associative principles of resemblance, contiguity, and cause and effect, the result is what Hume called the understanding. When it works capriciously, the result is fancy. Of course, the principles of association were as old as Aristotle, though Aristotle’s principles were not the same as Hume’s. Hobbes, too, had made use of them, though he believed that thought which was guided by desire or which exhibited a plan was more important. However, in Hume’s system for the first time they were looked upon as important *scientific* principles governing the working of the mind. This conception was taken up by Hartley in his theory of vibrations and developed into the associationist school of psychology.

Hume’s theory was also important in the history of psychology because it firmly established psychology as the science of the contents of consciousness. Although Descartes’s first certainty was rejected in relation to its content, what persisted was the assumption that a man

has some incorrigible sort of knowledge about his own mental states. Hume rejected Descartes’s search for simple natures, which appear to the mind as clear and distinct ideas, as the foundations of science. Instead, he postulated simple impressions of sense, perishing existences about which we can be certain provided that we make no inferences beyond them. Because Hume, like Locke, consistently confused psychology with epistemology, two parallel traditions developed from his work. On one hand, there was the search in epistemology for sense data which could provide an incorrigible basis for a system of knowledge; on the other hand, there was the development of introspective psychology whose task was envisaged as cataloguing the contents of the mind, analyzing them into simple units, and attempting generalizations about the links between these units which explained the generation of complex ideas and states.

*Body and mind.* Hume, understandably enough, had little to say about the relationship between mind and body. Body, according to his theory, stood for another bundle of impressions. He did not even connect the idea of self with impressions of bodily states, which might have been an obvious move if he had looked seriously for specific impressions, from which the idea of self is derived. In the Humean tradition William James, for instance, later suggested that the idea of self was intimately connected with impressions of breathing, cephalic movements, and the like. But Hume made no such suggestion. He noted the inexplicability of the fact that “the motion of our body follows upon the command of our will.” “Will,” he suggested, was another name for the strongest motive (compare to Hobbes’s account). But we simply have to accept these *de facto* connections between events. To speculate further would be to postulate occult causes and thus to sin against both Newtonian methodology and the way of ideas.

KANT. It would be very difficult to sketch the contribution of Kant (1724–1804) to psychology within the framework previously used, partly because he made very little direct and explicit contribution to psychology and partly because his Copernican revolution in philosophy involved a radical reformulation of questions asked under such a framework. Furthermore, though Kant’s concept of mind may, in fact, be extremely important insofar as it delimits the sphere of empirical psychology, those who developed empirical psychology in fact paid little heed to the implications of Kant’s position. Perhaps that was a pity, for Kant made a sustained effort to separate epistemology from empirical psychology, and until these two are clearly distinguished, there will continue to be confu-

sion in this area, as is demonstrated in the genetic psychology of Jean Piaget. Nevertheless, Kant's influence on psychology was largely negative and indirect; thus, only a short exposition will be given of those parts of his critical philosophy which seem relevant to psychology.

First and foremost, Kant rejected the notion of the empiricists that what is called mind could be explained as the product of ideas arising from experience and systematizing themselves according to laws of association. Kant maintained that the mind must be regarded as a structure regulated by principles of its own activity. These principles could not be arrived at empirically, for they were presupposed by any empirical investigation, including psychology. They could be arrived at only by critical philosophy, which asked the question "What must be presupposed for our experience to be possible?"

Kant was particularly interested in two realms of experience—Newtonian science and the autonomous morality of thinkers of the French Revolution. Kant attempted to reconcile the rationalism of Christian Wolff and Leibniz with the empiricist position of Hume by postulating an active mind whose nature was to impose a structure on experience to make it intelligible. This structure was composed of the categories used by scientists, such as substance, cause and effect, and continuity, which Hume had assigned to the imagination; Kant attributed the structure to reason, which synthesizes the data of sense. The content is provided by the senses, but the form is provided by reason. Thus, what we call nature is in part the work of mind. It is composed ultimately of things-in-themselves, whose real nature must be forever unknowable. We, too, must exist as noumenal selves, as things-in-ourselves. Of course, Hume was right in maintaining that we have no impressions of such selves. At best, we have intimations of such selves behind the appearances in our moral experience as active rational beings.

Human beings have empirical selves insofar as they have bodies and psychic functions—for example, sensation, imagery, feeling, purposes—which depend on embodiment. Such selves can be known by inner sense, and their manifestations can be investigated empirically; Kant called such a study anthropology. Kant made his mark on the history of introspective psychology by imposing on these phenomena the tripartite division—knowing, feeling, and willing—worked out in his *Critique of Judgment*. But he did not note anything particularly novel about the phenomena thus investigated, although he did declare that such investigations could never be properly scientific. He was convinced that science involved quantification and that since the phenomena

studied by anthropology could not be subsumed under mathematically expressed laws, psychology could at best be a collection of descriptive material classified under the headings that he suggested. Thus, Kant's extrapolation of Newtonian physics as the paradigm of all sciences had the negative effect of making it incumbent on those who wanted to develop psychology as a science to attempt the quantification of the phenomena to be studied. The result was Gustav Theodor Fechner's psychophysics, Johann Friedrich Herbart's attempt at mathematical laws of consciousness, and countless other premature attempts at quantification.

Another result of Kant's analysis was an increase of interest in the problems connected with the self. The controversy about the existence of a pure self and whether it was a proper object of study occupied most thinkers during the nineteenth century. Of much more importance for psychology, however, was Kant's doctrine that there can be no science of human actions, though its importance has seldom been recognized by those who are committed to empirical psychology. Human actions are the product of human reason, deliberation, and choice, and Kant held that insofar as a man's reason is involved, his behavior is not explicable in terms of the mechanical laws of nature. He acts freely and is determined only by rational laws of his own creation. This was similar to Spinoza's doctrine of freedom and activity. It raises all sorts of problems about the relationship between reason and emotion and between mind and body, problems that Kant did not seriously tackle. His concept of a rational being as a noumenon which was somehow related to a phenomenal embodied self was a metaphysical model that dramatized difficulties connected with the mechanical explanation of thought and rational action which Descartes had used a different model to depict. Kant laid more stress on the concept of will and rational action than did Descartes, but both men picked out a crucial problem for the development of psychology to which no satisfactory answer has yet been given.

## TRANSITION FROM PHILOSOPHY TO SCIENCE

The history of psychology as thus far reviewed is in the main a history of the philosophy of mind, and the issues discussed have been mainly philosophical issues. The rest of the history, however, will be concerned with the slow but progressive disentanglement of psychology as an empirical science from philosophical speculation.

Although it is possible to consider Aristotle's *De Anima* as the transition from presystematic to systematic

psychology, the transition from philosophy to empirical science cannot be pinpointed so precisely. This was not so much a transition as a process of differentiation. Indeed, it began with Aristotle, but it becomes unmistakable in the psychologies of Descartes and Hobbes, both of whom were affected by the impact of Galileo's physics. Both framed hypotheses about the physical and physiological mechanisms of consciousness and behavior that were in principle testable by observation and experiment. From Descartes and Hobbes the main line of development in empirical psychology was through the British empiricists Locke, Berkeley, and Hume.

#### EIGHTEENTH-CENTURY BRITISH PSYCHOLOGY.

Locke's new way of ideas laid the foundations for the twin doctrines of sensationism and associationism. The theory was that the mind is composed only of sensations and mental images (mental images being faint copies of sensations), that all complex percepts or ideas are formed through association, and that all trains of thought arise through association. Locke's analysis of mind was not so simple as that. He included ideas of reflection, abstract ideas, and the self, or possessor of sensations and ideas. Berkeley contested the existence of abstract ideas and furthered the development of associationism by giving an associationist explanation of the perception of the third dimension of space—another hypothesis that was to become the subject of experimental study. Hume further refined sensationism by eliminating the self on the basis of the negative result of his attempt to observe it by introspection. The next important step was taken by Hartley, who proposed a neural basis of conscious processes. His hypotheses, too, could in principle be tested by observation and experiment. Further refinements and elaborations of associationism are to be found in the works of James Mill, J. S. Mill, Thomas Brown, and Alexander Bain. The associationist doctrines spread to the Continent and as experimental psychology later returned to England and went to the United States.

A second major influence on the advance of psychology toward the status of an empirical science was provided by the biological sciences, notably in the evolutionary doctrine of Darwin. This influence was later to prove one of the causes of the disruption of associationist psychology.

*Hartley.* While David Hartley (1705–1757) was practicing medicine, he made many observations of psychological interest and wrote his major opus, *Observations on Man, His Frame, His Duty, and His Expectations* (1749). It was a thoroughgoing attempt to provide a neurophysio-

logical basis for the mental processes of sensation, imagery, and association. Influenced by Newton's *Opticks*, he proposed an explanation of conscious experience and association in terms of vibrations transmitted through nerves, which were conceived of as solid fibers, thus breaking from the earlier conception of nerves as hollow tubes for the conduction of the animal spirits. For every kind of sensation there are different kinds of vibrations or vibrations differently located; corresponding to images or memories, there are vibratiuncles, miniature vibrations that can persist after the larger vibrations have subsided and which form the physical substratum of memory. The associative processes occur by virtue of the fact that if two stimuli occur simultaneously and produce two corresponding vibrations in two regions of the brain—say, vibration *A* arising from a visual stimulus and vibration *B* arising from an auditory stimulus—the repetition of only the visual stimulus producing vibration *A* will arouse vibration *B* in the absence of the original stimulus that produced *B*. This is a simple translation into neurophysiological terms of the traditional principle of association of ideas, explaining, for example, the association of thunder with lightning. Hartley further advanced associationist theory by suggesting ways in which some of the several special laws of association—contiguity in space, contiguity in time, contrast, and similarity—could be reduced to the single law of association by temporal contiguity. He also offered a more detailed account than had yet been given, in terms of association, of the formation of general ideas.

*Brown.* As professor of moral philosophy in Edinburgh, Thomas Brown (1778–1820) delivered a series of lectures subsequently published under the title *Lectures on the Philosophy of the Human Mind*. Though not himself an associationist, he made very important contributions to the theory of association, which he preferred to describe as suggestion. Two of his ideas were of especial importance. First, he distinguished between simple suggestion, which is association in the commonly accepted sense, and relative suggestion, which is not in any sense an associative process but is a process that was later to be described by Charles Spearman as the “education of relations.” Second, Brown formulated the secondary laws of association—the principles of recency, frequency, duration, liveliness, and so on. These were later to become the subject of innumerable experimental studies.

#### NINETEENTH-CENTURY BRITISH PSYCHOLOGY.

Brown's philosophy was severely criticized by Sir William Hamilton (1788–1856) in his *Discussions on Philosophy and Literature* (1852) and his *Lectures on Metaphysics and*

*Logic* (posthumously published in 1859–1860), but Brown was defended with no less force by J. S. Mill in *An Examination of Sir William Hamilton's Philosophy* (1865). Hamilton, who was professor of logic and metaphysics at Edinburgh from 1836 until his death, had been greatly attracted by German philosophy and contributed to the rise of the British idealistic school of philosophy later to be represented by T. H. Green and F. H. Bradley. This school, deriving its inspiration from the intellectualist and idealist thought of G. W. F. Hegel and other Continental philosophers, had no common ground with the mechanistic empiricist and physiological approach of the British psychologists, but in its criticism contributed to the refinement, as well as the demise, of associationism. It was Bradley who, in attacking the atomistic features in associationism, phrased the dictum “Association marries only Universals.” This theme was to be developed in an original way in G. F. Stout's doctrines of noetic synthesis and relative suggestion.

**James Mill and John Stuart Mill.** Associationism reached its zenith in the work of James Mill (1773–1836). An economist and historian rather than a philosopher or psychologist, he learned his philosophy—hedonistic utilitarianism—from Hartley. His psychology, however, was a refinement of Hartley's and his analysis of mind was much more acute. *The Analysis of the Human Mind* appeared in 1829. Mental life was reduced to sensory elements, and the development of complex ideas was explained by the principle of association. Mill gave a clearer account than had Hartley of the way in which the several laws of association could be reduced to the single law of contiguity. He refined previous accounts of emotional experience in terms of sensations. Like Hartley, he attempted to apply the principles of associationism to the explanation of the complex phenomena of conscience and religion.

John Stuart Mill (1806–1873), his son, was a more subtle and acute philosopher than his father. He was certainly more disposed to take seriously any objection to a theory he wished to defend. In his rational and reasonable way he was inclined to make concessions that resulted in his rejecting the original theory. He sacrificed simple hedonism by conceding that pleasures might differ in quality. He gave up associationism by introducing the concept of mental chemistry—the idea that mental compounds, like chemical compounds, might exhibit properties not deducible from the properties of the elements. This breach in the associationist defenses was to be widened later by doctrines of creative synthesis and Gestalt qualities and the biological concept of emergent

evolution—ideas all at variance with pure associationist doctrines. J. S. Mill was less concerned with sensationism as a psychological doctrine than with its philosophical counterpart, phenomenalism—the description of material things and the physical world in terms of sense data or “permanent possibilities of sensation.”

**Bain.** Though in the associationist tradition, Alexander Bain (1818–1903) was less interested in the philosophy of mind than in psychology as an empirical science. He was emphatic in his demand that psychology should be cleared of metaphysics. His *Manual of Mental and Moral Science* (1868) was virtually a textbook of empirical psychology. It was a condensation of his two major works, *The Senses and the Intellect* (1855; rev. ed., 1894) and *The Emotions and the Will* (1859). He was thoroughgoing in his insistence on the need for a physiological basis for psychology not merely in general terms but in terms of known physiological facts, about which he made it his business to be well informed. As far as this implied a philosophy of mind, it found expression in his formulation of the principle of psychophysical parallelism. Especially important were his accounts of habit formation and learning. His treatment of habit was in large measure the inspiration behind the eloquent chapter on this topic in William James's *Principles of Psychology*. E. L. Thorndike and other “learning theorists” owe to Bain the first clear formulation of the law of effect, the principle that responses are ingrained by the reward of pleasure. Even his sillier theories contributed to enlightenment. One of the silliest theories in the history of psychology—that maternal love is based on the pleasurable tactile sensations experienced from contact with a baby—foreshadows the subtler theories of Freud concerning erogenous zones in the body and, more remotely, the “releaser mechanisms” of the ethologists. Bain's associationism was not an ideology. It was merely that he had assimilated the dominant features of the current psychological climate of opinion.

Two other developments were to complete the transformation of psychology from a branch of philosophy into an empirical science: (1) the impact of the theory of evolution and (2) the establishment of laboratories for experimental psychology. The theory of evolution had its origin in England in the work of Darwin; the idea of laboratories for experimental psychology came chiefly from the Continent.

**EVOLUTIONARY PSYCHOLOGY.** Darwin's theory of evolution as set out in his *Origin of Species* (1859) was a very large theory, but it was a scientific, not a philosoph-

ical, theory. It was supported by an enormous body of empirical observations. Theories of evolution date back to antiquity. Charles Darwin's grandfather Erasmus Darwin had adumbrated a Lamarckian theory of evolution. Alfred Russel Wallace anticipated Darwin's theory by a few months. Herbert Spencer (1820–1903), who had propounded philosophical and psychological theories of evolution for some years before the appearance of the *Origin of Species*, was accordingly well placed to capitalize on Darwinism in the development of his own ambitious “synthetic philosophy.”

Darwin (1809–1882) himself wrote on distinctively psychological topics. His *Descent of Man* (1871) discusses the similarities between the mental processes of man and of animals. His work *Expression of Emotions in Man and Animals* gives an evolutionary interpretation of changes in features and postures and assigns biological utility to these changes. The evolutionary approach stimulated many studies by amateur and professional naturalists. G. J. Romanes (1848–1894) collected evidence for the continuity of development from the animal to the human mind, and Sir John Lubbock (1834–1913) was among the first to use laboratory techniques in the study of insects. Laboratory studies like these were to be developed later on a grand scale by such American comparative psychologists as E. L. Thorndike and R. M. Yerkes.

**Galton.** Sir Francis Galton (1822–1911), the versatile cousin of Charles Darwin, contributed to meteorology, anthropology, anthropometry, and psychology and to the development of statistical and other metric methods in psychology. Among his major interests was the inheritance of mental characteristics, for the study of which he devised ingenious methods. He stressed heredity as a determinant of mental life and behavior. His records of the behavior of twins are reminiscent of the Leibnizian concept of a preestablished harmony. According to his records, twins can behave exactly like two clocks each causally insulated from environmental influences and from each other, behaving similarly and thinking in unison almost entirely in consequence of the similarities of their innate constitution. His major psychological works were *Hereditary Genius* (1869) and *Inquiries into Human Faculty* (1883). He set up the first two English psychological laboratories—the first at the International Health Exhibition of 1884 and the second in the South Kensington Museum. He pioneered the application of physical and psychometric tests in schools.

**Ward and Stout.** Philosophical psychology was to feel the impact of the new biological approach. James Ward's revolutionary article on psychology in the ninth edition

of the *Encyclopaedia Britannica* (1886) mounted a devastating attack upon associationism, recasting psychology in terms of a “psychoplasm,” or “presentational continuum,” which, like bodily tissues, undergoes progressive differentiation and integration. Ward's distinguished pupil Stout wrote *Manual of Psychology*, a standard text for some three decades, in 1898. This was described as being written from a genetic point of view; thereafter, almost every textbook of psychology had a biological orientation.

**EMPIRICISM IN EUROPE.** The empiricist philosophy was introduced into France by *littérateurs* and essayists like Voltaire and Denis Diderot, not by philosophers or psychologists. Voltaire had lived in England from 1726 to 1729, and so was in a position to introduce British ways of thought in philosophy into the intellectual life of France. Diderot had a clearer understanding of British empirical psychology. He particularly interested himself in the mental life of persons deprived of one sense—for example, sight.

The first of the French empiricist philosophers to contribute to sensationism was Étienne Bonnot de Condillac (1715–1780). Diderot had been concerned with the mental life of persons deprived of one sense; Condillac started from the imaginary case of a person deprived of all senses except one. He took the case of a statue endowed only with the sense of smell, selecting smell because of its relative simplicity. From this he proceeded to add other senses and to explain in sensationist terms attention, memory, imagination, and reason. He attached no importance to association. He believed that the experience of one sensation after another is *ipso facto* a comparison of the two and that the occurrence of the unpleasant sensation constitutes the will to terminate the sensation. Condillac's sensationism was perhaps the simplest and most elegant form of the doctrine in the history of psychology. His views are set out in the *Traité des sensations* (1754).

Claude-Adrien Helvétius (1715–1771), author of a volume of essays titled *De l'esprit* (1758), was a minor social and political philosopher who seized upon Locke's empiricism and concept of the tabula rasa to defend an extreme doctrine concerning the equality and perfectibility of men. His basic thesis was that all differences between men are due to differences in experience and education. All error was due to passion or ignorance.

The doctrines that Helvétius derived from Locke were to return to England in the works of William Godwin, especially in his *Political Justice* (1793). Like

Helvétius, Godwin taught that all men are equal at birth and that their subsequent differences were due to experience and education. Voluntary actions originate in opinions, which can be changed by rational persuasion. Vice is error, which can be corrected. In Helvétius and in Godwin the association of empirical philosophy with an intellectualist hedonism is displayed in its most extreme form.

Through Condillac the influence of Locke spread to Italy and Switzerland. In Italy this influence is to be seen in the teachings of several all-but-forgotten writers. In Switzerland, Charles Bonnet (1720–1793) of Geneva was the outstanding figure in empirical philosophy. His chief work in psychology was the *Essai analytique sur les facultés de l'âme* (1760). Although he followed Condillac for the most part, Bonnet differed chiefly in the importance he attached to physiological explanations.

**GERMAN PSYCHOLOGY AND EXPERIMENTATION.** Throughout the seventeenth, eighteenth, and early nineteenth centuries German psychology was dominated by the philosophical doctrines of Leibniz, Kant, and Hegel, each of whom contributed to a rationalist idealism very unfavorable to the development of psychology as a science.

*Hegel.* Georg Wilhelm Friedrich Hegel (1770–1831) has received scant attention in the histories of psychology, understandably so since his form of rationalism is the most extreme antithesis to the empiricist philosophy that had favored the development of psychology as an empirical science. He is, however, not without importance in the history of psychology.

One of Hegel's theses was that it is a mistake to suppose that complex phenomena are explained only by reference to simpler phenomena, that we can, for example, understand religion in its developed form by the study of cults of primitive people or that we can understand man only through the study of lower forms of animal life. In this he challenged what had long been and still is a basic principle of comparative psychology, but Hegel's thesis survives in the view of psychologists who hold that the proper study of humankind is man and that we should begin with civilized man in advanced societies. It lives on in the contention of Freudian psychologists that the evidence for infantile sexuality can be appreciated only in the light of adult sexual behavior.

Equally important for psychology is the Hegelian dialectical progression—thesis, antithesis, synthesis. When this progression is stated as an empirical observation of movements of thought and action, not as a metaphysical principle or a principle of logic, it illuminates

many sequences in the history of politics, philosophy, and science. A dialectical progression is illustrated in the fate of Hegel's own philosophy. Its influence in Germany was short-lived. His rationalistic thesis issued in an empiricist antithesis, Wundtian experimental psychology. The dialectical progression is illustrated by the British vogue for Hegelianism among philosophers who found in it an antithesis with which to confront the prevailing empiricist philosophy and psychology. The progression is illustrated by the sequence from Hegel's idealist thesis to the antithesis of dialectical materialism that was to become a central tenet of communist philosophy. Although it provides no comprehensive philosophy of history, the concept of dialectical progression affords a rather more subtle and articulate account of historical movements than conventional, commonsense accounts in terms of "the swing of the pendulum."

Hegel's doctrines were associated with, and conferred philosophical status upon, a widespread romantic and mystical philosophy of nature according to which everything in nature had some spiritual and symbolical significance. The influence of this philosophy of nature persisted far into the nineteenth century and in the biological sciences favored vitalistic, as opposed to mechanistic, accounts of mind, body, and nature. Psychologists divided progressively into two groups. The first comprised the philosophers—that is, those who primarily taught philosophy and whose philosophy of mind contained much metaphysics. The second group consisted of natural scientists whose approach was from mathematics, physics, and the biological sciences. The distinction is not sharp, since romanticism and metaphysics were in the air that every German student, even students of the natural sciences, breathed.

The first steps in the transition from the philosophy of mind to scientific psychology were taken when Kant challenged psychologists to show that their subject could claim scientific status. This challenge was taken up by Herbart, Ernst Heinrich Weber, and Fechner. That it could be an experimental science was argued by Weber, Fechner, Johannes Müller, Hermann von Helmholtz, and others. Wundt finally established it as a science that required a distinctive kind of laboratory.

*Herbart.* Johann Friedrich Herbart (1776–1841) set out to establish a basis for psychology other than that of the prevailing "faculty" psychology associated with Christian Wolff (1679–1754), a disciple of Leibniz and precursor of Kant who was much less distinguished than either. Herbart tried to show that the laws describing mental process could be put into precise mathematical form.

Herbart's first achievement was the grafting of associationism onto a rationalist metaphysical root. The soul was retained, serving the traditional function of giving unity to the mind, but the data of empirical psychology were, as in associationism, sensations and ideas. In Herbart's system ideas were not just passively associated. They interacted by attractions and repulsions in accordance with which they were drawn into or forced out of consciousness. The behavior of ideas in Herbart's psychology resembles that of the "reals" in his pluralistic metaphysics. Two "reals"—for instance, *A* and *B*—differing in quality, tend to disturb each other because of their difference, but each also tends to preserve itself by resisting the disturbing effect of the other. This principle of self-preservation is reminiscent of the Spinozistic doctrine that "everything that is in itself endeavors to persist in its own being" and, when applied in Herbart's psychology, foreshadows the concept of homeostasis that was to be current in psychology a century later.

Herbart's account of the way in which ideas enter consciousness and are expelled from it represents a phase in the history of the theory of the unconscious midway between Leibniz and Freud; his concept of the apperceptive mass, a system of ideas bound together by mutual attraction, was still current when psychoanalytic writers were developing the concept of a mental complex. Herbart's metaphysics and mathematics were to be forgotten, and he did not contribute directly to the development of psychology as an experimental science. His most lasting influence was in the field of educational psychology, chiefly in the application of his theory of apperception to the process of learning.

**Lotze.** Rudolf Hermann Lotze (1817–1881) succeeded Herbart in the chair of philosophy at Göttingen. His most influential work was his *Medizinische Psychologie* (1852), the first systematic work on physiological psychology and one of the very few written by an author qualified in both physiology and philosophy. Against the then prevailing view he defended the thesis that every mental phenomenon has its physiological counterpart and that the laws which apply to inorganic matter also apply to organic matter. Final causes, vital and mental forces, and the soul itself can act only through mechanical causation. He insisted, however, that physiology alone cannot explain mental phenomena. Lotze is best known in psychology for his doctrine of local signs, a contribution to the theory of space perception.

**Weber and Fechner.** Experimentation and the use of quantitative methods in psychology were greatly advanced by Ernst Heinrich Weber (1795–1878) and

Gustav Theodor Fechner (1801–1887), who were colleagues in the University of Leipzig and who both taught Lotze.

Weber taught anatomy and physiology. His early work *De Tactu* (1834) reported studies demonstrating the difference between muscle sense and touch. These studies were extended to pain, pressure, and temperature, through which emerged the concept of thresholds and the famous law that has come to be called Weber's law. This states that the smallest increment in a stimulus required to produce a difference in the sensation experienced is not an absolute amount but is relative to the magnitude of the stimulus in question. Like most German scientists of his time, Weber was to some degree under the spell of the current metaphysics and the romantic philosophy of nature, but neither of these influenced his experimental studies. His metaphysics and his science were kept apart.

With Fechner the case was different. Fechner's intellectual life was a pilgrimage from physics and chemistry, through physiology and medicine, to metaphysics and mysticism. From an early age he had been preoccupied with the problem of the relation between matter and spirit. He was attracted to a form of panpsychism according to which not only man and the lower animals have consciousness but also the earth and the other planets—indeed, all material things. In this view all souls are parts of the soul of the universe.

Fechner concluded, on the obscurest of grounds, that the mystery of the relation between mind and body would be resolved by ascertaining the quantitative relations between stimuli and sensations. He suggested that Weber's law could be put into a quantitative form. Weber's law thus became the Weber-Fechner law, according to which the relation between stimulus and sensation is expressed in the formula  $S = k \log R$  where *S* is the experienced intensity, *R* is the physical intensity, and *k* is a constant for the particular sense in question. For the verification of this law Fechner designed what are known as the psychophysical methods. These methods have been used in the most tedious of laboratory exercises to which many generations of students of experimental psychology have since been subjected, and the published results of these exercises are among the most tedious controversies in the history of science. But the possibility of experiment and measurement in psychology was established—paradoxically, by a metaphysical mystic. The metaphysics and the mysticism were soon forgotten, but the exercises live on.

**Beneke.** Friedrich Eduard Beneke (1798–1854), a contemporary of Herbart, was another philosopher who contributed to the foundation of a science of empirical psychology, which, he claimed, was the basis of all philosophy. Like Herbart, he set out to provide a basis for psychology other than that of a doctrine of faculties, and like Herbart, he stressed the activity of the mind. Among his works on psychology are *Lehrbuch der Psychologie* (1832) and *Die neue Psychologie* (1845). Because of his rejection of the prevailing Hegelian philosophy of the Absolute, Beneke was dismissed from his post in the University of Berlin, but after Hegel's death he was reinstated. His best-known contribution to psychological theory was his doctrine of mental, as contrasted with physiological, traces for the explanation of the facts of memory. This doctrine was later to be developed in Great Britain by Stout.

**Müller.** Johannes Müller (1801–1858) was a contemporary of Beneke at Berlin. He was the first to hold the title of professor of physiology. (Hitherto, the subject had been taught as a branch of medicine.) He had been under the influence of the prevailing philosophy of nature but contributed to the clarification of the concepts of mind, body, and nature by distinguishing the mental principle, which is restricted in its operation to the nervous system, from the vital principle, which is diffused throughout the organism. He was also preoccupied with the opposition between nativistic and empiricist explanations of space perception as represented, respectively, in the doctrines of Kant and Herbart. Müller reformulated the issue in terms that made it possible to submit the question to experimental tests. He also formulated the theory of specific energies in the nervous system—the hypothesis that the sensory qualities are generated by specific activities of the organs of sense or by specific differentiation in corresponding realities in the brain.

**Helmholtz.** Hermann von Helmholtz (1821–1894), Müller's distinguished pupil, is acknowledged to be the most outstanding of the physicist-physiologists who have contributed to the development of experimental psychology. In the range of his pioneering studies he has been compared with Francis Galton. His publications were more numerous than Galton's, and his investigations were carried further. He was the first to make a realistic calculation of the speed of nervous impulses, which are important, among other things, in the study of reaction times. He developed Müller's doctrine of specific energies and Thomas Young's three-color theory of vision.

Helmholtz's *Handbuch der physiologischen Optik*, published in three volumes (1856–1866), remained an authoritative text for many decades, although it was not

translated into English until 1924–1925. No less outstanding were his contributions to the theory of hearing and the related subjects of phonetics and music. He was essentially a scientist with little interest in philosophy and still less patience with transcendentalism. There is, however, much in his writings of philosophical interest—for example, his puzzling concept of unconscious inference in perceptual judgments. His discussions of the principle of the conservation of energy are important in the history and philosophy of science.

**Wundt.** The last phase in the transition of psychology from a branch of philosophy to psychology as an independent empirical science is conveniently dated as beginning in 1879, when Wilhelm Wundt (1832–1920) established the first psychological laboratory. Wundt's chief claim to a place in the history of philosophy arises from the conceptual system in terms of which he interpreted the experimental data from his own and other laboratories. His philosophy of mind deviated from the simpler forms of atomistic sensationism in that the ultimate elements of mind were, according to him, of two kinds, sensations and feelings. He and his disciples devoted much energy and skill to defining the differences between sensations and feelings and to elucidating his curious tridimensional theory of feeling, but the general program was to analyze experience into its elements, to define the fundamental attributes of these elements, and to formulate the laws in accordance with which these elements are combined. The account leaned heavily on the principle of association but deviated from traditional associationist doctrines in introducing a concept of creative synthesis. This concept was a variant of the concept of apperception and embodied a theory of attention. It had some points in common with J. S. Mill's conception of mental chemistry and in some degree foreshadowed later theories of emergent properties and the doctrine of the Gestalt psychologists that a complex experience is more than the sum of its parts. His most influential work was *Grundzüge der physiologischen Psychologie* (1873). In later years he published two works that contributed to the incursion of psychology into sociology and anthropology.

**Ebbinghaus and Külpe.** Among other outstanding experimental psychologists were two of Wundt's pupils, Hermann Ebbinghaus (1850–1909) and Oswald Külpe (1862–1915). Wundt's laboratory research had been chiefly concerned with sensation and perception and with relatively simple processes of reaction and association. Ebbinghaus and Külpe extended the experimental



method into the study of the higher and more complex functions of memory and the processes of thinking.

In a monumental work, *Über das Gedächtnis* (1885), Ebbinghaus published the results of what has been described by J. C. Flügel in his *A Hundred Years of Psychology* as “the most brilliant single investigation that has ever been made in experimental psychology.” Ebbinghaus’s outstanding achievement was to extend the experimental method to the “higher thought processes.” He was the first to establish quantitative laws concerning the process of memorization. In 1894 he succeeded Theodor Lipps, a pupil of Wundt’s most widely known for his studies in psychological aesthetics, in the chair of psychology at Breslau. There Ebbinghaus pioneered in the study of intelligence and devised the completion test, which remains an important component of intelligence tests.

Külpe directed the laboratory at Würzburg, which achieved great fame through its investigations of willing and judging. Through the discovery of imageless thoughts these studies contributed both to the breakdown of sensationism and, in consequence of the inconclusive disputes this discovery provoked, to the behaviorist revolt against introspective methods. At Würzburg as at Leipzig confusion arose through the interpretation of experimental data in terms of implicit philosophical concepts and assumptions, and the conclusions drawn have had to wait for review in the light of further clarification of the distinction between empirical psychology and the philosophy of mind.

**THE SHIFT TO THE UNITED STATES.** In the age of Wundt, psychology was a Germanic science, and Germany was the heart of the empire. Mainly through Wundt’s influence upon those who came to Leipzig from the United States, psychology became an American science with the United States as the new seat of dominance. Among those who studied abroad and then returned to America were Stanley Hall, who established the first American psychological laboratory at Johns Hopkins in 1888; J. McKeen Cattell, who after several years as assistant to Wundt founded the laboratory at Pennsylvania; and Hugo Münsterberg, who, having established a laboratory at Freiburg, was invited by William James to Harvard in 1892. In the same year E. W. Scripture took charge of the laboratory at Yale. By 1897 there were fifteen psychological laboratories in the United States, and by the end of the century there were twenty-six, all based, to begin with, on the laboratory in Leipzig. Most of Wundt’s American pupils, however, were soon to deviate from the

German pattern and to open up approaches characteristically American—allergic to philosophical speculation, distrustful of introspective methods, and much concerned with the practical applications of their science. Hall became famous for his studies of adolescence. Cattell, more influenced by Galton than by Wundt, concentrated on the measurement of individual differences. Münsterberg’s interest turned to applications of psychology to industry and criminology. The mantle of Wundt fell upon E. B. Titchener, an Englishman from Oxford who after his studies at Leipzig went to the United States to develop experimental psychology at Cornell.

**THE ESTABLISHED ORDER OF 1900.** Wundtian psychology was one important form of and ingredient in what has been called the established order of 1900, against which many revolts were to be mounted. There were, in fact, at least two established orders, one in Britain, represented by Ward and Stout, and the other in the United States, represented by Titchener. These were very different establishments, but they had in common a foundation in some form of body-mind dualism and the acceptance of the facts of consciousness, observed by introspection, as defining the subject matter of psychology.

**Ward.** James Ward (1843–1925) presented his own system as a sort of synthesis of the too objective thesis of Aristotle’s psychology and the too subjective antithesis of Descartes’s psychology. His basic conceptual framework was doubly tripartite. In his analysis of experience he distinguished the three modes of consciousness—cognition, feeling, and conation; his analysis of each kind of experience referred to a self or ego, an act or mental attitude, and a presentation (a mental object, sensation, or idea). The most interesting features of his system are contained in his detailed analysis of the phases of development from simple sensation to perception and from perception to the construction of a memory thread and an ideational tissue. Though qua psychologist Ward can be treated as a dualist, his background metaphysics was a variant of an idealistic monadology of the Leibnizian type.

**Stout.** G. F. Stout (1860–1944) developed Ward’s psychology in an individual way, creating an original and independent system. As a psychologist Stout, like Ward, developed a dualistic psychology, but as a philosopher he developed an original theory of mind, body, and nature.

**Titchener.** E. B. Titchener’s laboratory at Cornell was the temple of the Wundtian form of the established order, and Titchener (1867–1927) was its high priest. Here as elsewhere, however, empirical psychology continued to

be inextricably entangled with philosophy. Titchener's deviations from sensationist and associationist psychology were less fundamental than he himself believed. He was a dualist, and he confessed to a bias in favor of sensationism. He was reductionist in his treatment of conation. He differed from the classical sensationists in accepting feelings as basic elements; he also differed from them in the treatment of the elements as existences, as contrasted with meanings. He sought to explain complex mental states as arising from the synthesis of elements and thus to display the structure of the mind. Accordingly, he is described as a structuralist, as opposed to a functionalist, psychologist. His cardinal tenet, which was to become the major object of attack, was his thoroughgoing proclamation of introspection as *the* distinctive method of psychology. His two most important works were his *Lectures on the Psychology of Feeling and Attention* (1908), a detailed exposition of the thesis that "the system of psychology rests upon a threefold foundation: the doctrine of sensation and image, the elementary doctrine of feeling and the doctrine of attention," and the *Lectures on the Experimental Psychology of the Thought* (1909), an equally thoroughgoing examination of the claims for the discovery of an imageless thought element and a polemic against the doctrine of Franz Brentano and Stout that references to object is the criterion of mind.

**REVOLTS AND THE ERA OF THE SCHOOLS.** The established order of the United States and the established order of Britain were to become the objects of attack from four directions: (1) The behaviorist attack directed in the main against dualism, the concept of consciousness, and the reliance upon introspection; (2) the attack of the Gestalt psychologists against all forms of psychological atomism; (3) the psychoanalytic attack against the overemphasis on conscious processes and inadequate recognition of the unconscious mind; and (4) the attack of the hormic psychologists, which was directed against the intellectualism of traditional psychology—that is, the overemphasis on cognitive processes and the relative disregard for conation or purposiveness in the explanation of conscious experience and behavior.

In the four revolts the schools were all fighting on more than one front. Each was attacking traditional psychology, and each engaged in polemics with the other revolting schools. Confusion was increased by the fact that within each school there were conflicting factions and by the general failure to distinguish straight empirical issues from issues of philosophy and of linguistic usage.

**Behaviorism.** The conception of psychology as the study of behavior and as an essentially biological science dates back to Aristotle, but behaviorism as an ideology can be dated precisely. It began in 1914, when J. B. Watson (1878–1958) published *Behavior* while a professor at Johns Hopkins University.

This book was a protest and a revolt against dualism, the concept of consciousness, and *any* use of the introspective method in psychology. Psychology is to be the study of behavior by objective methods. It was a protest in defense of animal psychology, in which statements about the animal mind and the consciousness of animals must be pure guesswork, and it was a protest against the interminable and inconclusive disputes between introspective psychologists about the differentiation of sensations and feelings, the James-Lange theory of emotion, and imageless thought. It was also an attack on the traditional theory of consciousness in which some sort of mental stuff was thought to be the subject matter of psychology.

In *Behavior* and two other important books, *Psychology from the Standpoint of a Behaviorist* (1919) and *Behaviorism* (1924), Watson developed his distinctive account of all the major topics that constitute psychology. Like the structuralists he set out to exhibit complexes in terms of simple elements, complex responses to situations as derived from simple responses, native and acquired. The analysis of behavior was in terms of stimulus and response (an analysis to be elaborated later by E. C. Tolman in terms of intervening variables). Sensation and perception were described as responses to present stimuli and constellations of stimuli, memory and learning as responses to past stimuli and neural traces, feelings and emotions as types of sensorimotor responses, and thinking as subvocal verbal behavior. Introspection itself was redescribed as verbal behavior. In his system Watson included much that was irrelevant to the major principle—for example, a bias toward explanations in terms of environmental influence and a bias against explanations in terms of heredity. He had a special bias against the concept of purpose, though later behaviorists found no difficulty in assimilating purposive behavior as goal-directedness. His laws of conditioning were the old laws of association transformed into generalizations about bonds between simple reflexes instead of between simple ideas.

Watson's writings were naive and often confused, but his behaviorism sailed on the tides of the time. Behaviorism was inevitable. Watson's behaviorism was fortunate in that it was reinforced by the most important philo-

sophical movements of the period, positivism and physicalism. It was also reinforced by the logicians and the methodologists of the inductive school, who maintained that scientific laws state correlations between observables. Watson accordingly assumed that because mind was unobservable, it could not be discussed or referred to in science. When logicians later began to proclaim that scientific systems were hypothetico-deductive, such behaviorists as Tolman and C. L. Hull conceded the importance of unobservables in the form of intervening variables and hypothetical constructs. This return to the methodology of Galileo made any simple form of behaviorism difficult to maintain. Nevertheless, B. F. Skinner stuck to the old inductive concept of scientific method and proclaimed that his findings involved no theory. Behaviorism was further supported by a number of outstanding experimental psychologists—for example, K. S. Lashley and W. S. Hunter—sympathetic to Watson's approach.

Lashley was primarily a neurophysiologist who as a behaviorist was more sympathetic to the views of the Gestalt psychologists than to those of Watson. He contributed in an important way to the advance of knowledge concerning the localization of the higher functions in the cortex.

Hunter, a distinguished experimental psychologist, rallied to the support of behaviorism through an odd philosophical argument, based on a very naive form of realism, that consciousness or experience is merely a name applied to what other people call the environment. This argument is reminiscent of a characteristic doctrine of Ward and Stout that the subject matter of psychology comprises "the whole choir of heaven and earth" as it appears to the observer, a view later to be defended by the Gestalt psychologists in terms of the behavioral, as contrasted with the geographical, environment—another variant of the view that things as they appear are appropriate objects of psychological science.

As professor of psychology at the University of California, E. C. Tolman (1886–1959) developed an original system of purposive behaviorism that had perhaps much more in common with the psychology of McDougall than it had with the psychology of Watson. Watson was preoccupied with responses to stimuli. Tolman described Watson's behaviorism as molecular, for it was concerned mostly with physiological details; his own he described as molar, for it was concerned with external and integrated behavior and with emergent properties.

Clark L. Hull (1884–1952), professor at Yale, is known for his inventiveness and originality. His contribution to behaviorism reflects his own interest in

methodological studies and the concept of hypothetico-deductive systems. He constructed a miniature system of this type aimed at a rigorous ordering of some of the basic laws of behavior. His deductive dream and his attempt to develop a Galileo-like resolution of behavior into simple externally initiated movements bore a marked similarity to the mechanistic system of Hobbes.

Behaviorism is not strictly an arguable thesis; it is a pronouncement, a policy statement. The traditional psychologist declares, "I propose to study consciousness by introspection"; the behaviorist says, "I do not; I propose to study behavior by objective methods." The issue is almost as simple as that. There are, however, many arguable issues in particular systems of behaviorism. Reasons can be given for and against policy decisions. There are larger philosophical issues that cannot be evaded.

Roughly three types of behaviorism have emerged: a metaphysical type that says that consciousness does not exist; a methodological type which says that consciousness is not amenable to scientific procedures of investigation; and a radical analytic type, defended chiefly by philosophers, according to which mental facts can all be analyzed in terms of behavior and dispositions to behavior. In Watson's behaviorism and in many others these issues are confused. The behaviorists, no less than Titchener, confused questions of empirical fact with questions of philosophical analysis. It is not possible to know what an emotion is by the introspective observation of emotional states. A prior decision has to be made concerning what to observe, what is to count as an emotion. In the same way it is not possible to know what behavior is by the objective observation of behavior. A prior decision has to be taken about what to observe and about what is to count and what is not to count as an example of behavior. For example, before describing a movement of the body like raising an arm as signaling to a friend or testing the direction of the wind, a person must know what the agent had in mind. This inadequate attention to the question of what constitutes behavior was one of the major weaknesses of behaviorism. Behaviorism is no less riddled by interminable and inconclusive disputes than is introspective psychology. Nevertheless, it has contributed very effectively to the advance of psychology as a biological and an experimental science.

*Gestalt psychology.* The term *Gestalt psychology* applies primarily to a school of psychology pioneered by Max Wertheimer (1880–1943), Kurt Koffka (1886–1941), and Wolfgang Köhler (b. 1887). Their polemic was directed chiefly against the atomism of traditional psychology and of the established order. They opposed the

thesis that perceptual experience is to be explained by a bricks and mortar account of the combining of simple sensations. Their positive thesis was that what is experienced is always organized and consists of wholes which are greater than the sum of their parts. Like all revolutionaries, they exaggerated the difference between their own ideology and traditional doctrine. The fact with which they were concerned had preoccupied philosophers and psychologists from the beginnings of systematic thought. Aristotle's formal cause was a Gestalt concept, and Kant had grappled with the problem in his treatment of the categories; Ward and Stout had grappled with it in their accounts of the development of the perception of space, time, thinghood, and causality, and Mill had seen the problem when he wrote about mental chemistry. Christian von Ehrenfels (1859–1932), an Austrian philosophical psychologist, introduced the concept of form qualities. There were also contemporary psychologists—for example, Charles Spearman and Henry J. Watt—who were concerned with the concepts of Gestalt psychology in their own ways.

The outstanding contribution of the Gestalt psychologists was in the number, the variety, and the ingenuity of their experiments. Wertheimer's elegant experiments on the perception of movement were followed by no less elegant experiments by himself, his colleagues, and his disciples on the principles of organization in perceptual experience. In the earlier phases Gestalt psychology was as intellectualist as traditional psychology in its preoccupation with the cognitive experience of the normal adult human mind. Its interest extended, however, to child psychology in the studies by Koffka and to animal psychology in Köhler's studies of insight and learning in apes. Kurt Lewin (1890–1947) used Gestalt concepts in the study of problems of personality and of human motivation. The Gestalt psychologists were distinguished chiefly by their experimental inquiries, but in their writings there are many pronouncements relevant to the philosophy of mind.

The slogan "The whole is more than the sum of its parts" is a near tautology but a useful tautology. The increasing emphasis placed by Köhler and Lewin on field theory (the theory concerning properties of total fields of activity as contrasted to the properties of isolated units) has also contributed to the philosophy of science in its application to psychology.

The concept of isomorphism (the parallelism between phenomenal experience and neural processes) has given a new slant to the discussion of classical theories concerning the relations of body and mind.

The experimental findings of the Gestalt psychologists have been assimilated into empirical psychology. Its evaluation as a philosophy of nature, life, and mind must take into account not only its historical antecedents but also some less well known but important contemporary theories, such as those, for example, of Spearman and Watt.

*Alternatives to Gestalt psychology.* Charles Spearman (1863–1945) made two significant contributions to the development of psychology in the early decades of the twentieth century. The first was through the development of statistical methods in psychology. Building on the studies of Galton and Karl Pearson, he elaborated his two-factor theory for the analysis of human abilities. His second notable contribution was an attempt to formulate principles of cognition, which he believed to be as basic to psychology as Newton's laws had been basic to physics. It was an ambitious plan in which three noegenetic principles—the apprehension of experience, the eduction of relations, and the eduction of correlates—were set out as necessary and sufficient for the explanation of all the cognitive operations of the human mind. The principles of the eduction of relations had been anticipated by Brown's concept of relative suggestion, but in its detailed elaboration it covered most of the facts of cognitive experience studied by the Gestalt psychologists.

Henry J. Watt (1879–1925) enters the history of psychology through his experimental studies of judgment and the higher thought process at the Würzburg laboratory. After his return to Britain he spent the rest of his life at the University of Glasgow elaborating a comprehensive theory that was finally presented in his *Sensory Basis and Structure of Knowledge* (1925). It is a paradoxical fact that atomism, against which Gestalt psychology was directed, should have received its most precise and systematic formulation by a psychologist preoccupied with precisely the facts that Gestalt psychologists were concerned with.

Watt offered an ingenious alternative to Gestalt theory made possible by the sharp distinction he drew between sensationism and associationism, whereas Titchener had treated them as equivalent doctrines. Watt agreed that traditional psychology rested upon two postulates—(1) that the elements of mind are sensations and (2) that the compounds are produced by association. He not only accepted the first postulate, but he also refined it with great subtlety. He rejected the second postulate, replacing it with the doctrine that complex cognitive experiences arise through a distinct process of integration—a concept to which he gave a new definition and which he illustrated in great detail. Watt produced an

original account of the facts that had previously been interpreted in terms of Mill's mental chemistry, Wundt's creative synthesis, Spearman's noetic principles of education, and the principles of Gestalt psychology.

The Gestalt psychologists captured the headlines in the journals of psychology. For a time Spearman had a band of disciples, although Watt's book did not have a second edition. Spearman and Watt had the misfortune of attracting disciples who could neither advance their theory nor excite impassioned critics. Thus, both have been forgotten. Both, however, may be classed among the mute inglorious Miltons of psychology whose works may yet attract the attention of future historians of science.

The philosophy of nature, life, and mind of both Spearman and Watt were, though different from each other, both in the tradition of dualism. That of the Gestalt psychologists was rather different—a dualism of physics and phenomenology. A residual doubt remains. There would appear to be no empirical procedure for deciding between the doctrines of the Gestalt psychologists, of Spearman, and of Watt. The case may again be one in which a choice must be made on grounds of terminological convenience.

*Psychoanalysis and derivative schools.* The most important revolt against traditional psychology at the turn of the twentieth century was that of Freud and his disciples.

Sigmund Freud created an entirely new psychology—psychoanalysis. This is both a technique of psychotherapy and a body of theory providing a rationale for the technique. The theory developed into an overall account of nature, life, and mind. Freud's philosophy of nature was a conventional nineteenth-century mechanistic materialism predisposing him to an equally conventional preference for physiological explanations of the mind. Thus, it is even more remarkable that his most distinctive and revolutionary doctrines assumed the form of hypotheses to which mechanism and physiology are completely irrelevant.

Central in his system of psychology is the concept of the unconscious. Mind is divided into the conscious, the preconscious, and the unconscious. The conscious is the traditional, familiar, introspectable part of the mind—introspectable thoughts, feelings, and desires. The preconscious consists of all that is out of mind but which can be brought to mind at will or which readily returns to mind in accordance with the accepted laws of association. The unconscious, on the other hand, consists of ideas and wishes, especially wishes, which can be brought into con-

sciousness only by special techniques, of which psychoanalysis is said to be the most fundamental.

Freud's originality did not consist in the discovery of the unconscious, for others before him had hit on this notion, but in postulating that the mind worked in accordance with two different types of laws—those of the primary processes, which included unconscious processes, and those of the secondary processes of thought. The first were ruled by the pleasure principle, the second by the reality principle. The laws of the primary processes were principles of emotive congruence appropriate to wishes. Freud's great contribution to psychological theory lay in postulating these laws of primary processes to explain such phenomena as hysteria, dreams, parapraxes, and so on which were previously unexplained and among which no one had previously seen any connection.

There are some superficial resemblances between Freud's and Herbart's psychology, but these are only superficial. In Herbart's system the contents of the unconscious were ideas; in Freud's system they were mainly wishes. Herbart was concerned with the movement of ideas between consciousness and Freud's preconscious. He had no clear conception of the unconscious mind in Freud's sense. Herbart's explanation of the movements of ideas were formulated in terms of quasi-mechanical forces, efficient causes, whereas Freud's explanatory principles were, in effect, formulated in terms of a truncated type of teleological concept, the Freudian wish. Similarly, Freud's defense mechanisms—sublimation, projection, reaction formation, and the like—were quite unmechanical mechanisms. They were goal-directed procedures for protecting the conscious mind against the unwelcome wishes and ideas that had been repressed.

From first to last Freud was concerned with mental conflict, the conflict between opposing motives. At the beginning he emphasized the conflict between primitive instinctive impulses, mainly sexual, and the need to conform to the rules and norms of society. The emphasis later shifted to the conflict between the life and death wishes. At first the world was astounded and shocked by Freud's theories about sex, especially by his account of infantile sexuality. So prominent was sex in his system that a Freudian explanation of any form of behavior came to be generally thought of as an explanation by reference to unconscious sexual desires. His generalized concept of sex was that all pleasure is essentially the pleasure of sexual experience, including the satisfaction of defecation (anal eroticism) and the satisfaction of sucking and feeding (oral eroticism), as well as the satisfaction derived

from the genital organs (genital eroticism). This general theory of affective experience makes the thesis of infantile sexuality almost tautological. More significant empirically was the thesis of the universality of the Oedipus complex—the thesis that every male child unconsciously wishes to kill his father and have sexual relations with his mother (female children have an Electra complex—the unconscious wish to replace the mother in her relation to the father). These unconscious desires are obvious sources of the conflicts that issue in neuroses and other forms of aberrant behavior.

In Freud's later writings the emphasis was transferred to the conflict between the life-promoting instincts and the desire for death—Eros and Thanatos. When directed outward, the death wish is a source of violence and destruction; when directed inward, it results in suicidal behavior. The concept of the death wish was, however, further generalized. It covered not only the desire to kill and to be destroyed but also the desire to inflict pain and to suffer pain. Thus the odd phenomena of sadism and masochism are explained. As he often did, Freud attempted to reinforce limited hypotheses by highly general theories. The hypothesis of the death wish was based upon the general theory that in all the processes of nature there is a tendency for animated matter to revert to an inorganic state. Slightly less generalized was the theory that all responses to stimuli by an organism were directed to the removal of the stimulus and are thus consummated in unconsciousness, in sleep or death. These speculations were disturbing to his disciples, who felt an obligation to defend them, since these ideas were all but demonstrably mistaken and on the face of it inconsistent with Freud's more basic hedonistic account of human motivation. They were not at all essential to his general theory.

To this phase of Freud's speculations belongs the doctrine that the total personality is organized on three levels—the id, the ego, and the superego. The id consists of the totality of primitive instinctive impulses, and the ego contains the conscious motives. The concept of the superego is the most interesting and original feature of this hierarchy. Although it was often described as the primitive unconscious conscience, Freud explained it as an introjected image of the parent that continued to issue commands and to administer punishment when those commands were disobeyed. Not a few of Freud's disciples have treated the superego as the source of conscience as traditionally conceived and believe it is the explanation of action that accords with moral principles. This, however, was not Freud's view. He was himself a man of great integrity with very definite ethical principles. These prin-

ciples were not derived from his own superego but are to be explained in terms of the distinction between the pleasure principle and the reality principle. Action in accordance with the pleasure principle is directed to immediate pleasure regardless of consequences; action in accordance with the reality principle is directed to maximizing pleasure in the long run. This may be little more than a terminological variation on traditional hedonism, but as is often the case, terminological innovation can contribute to enlightenment.

By 1950 Freudian theory was the dominating influence in psychology. Neither the technique of psychoanalysis nor the supporting theory has received scientific validation, but no theory of human motivation and no form of psychotherapy can ignore the theories and practice of Freud. Freud himself protested that psychoanalysis does not attempt to explain everything, but in the human and social sciences there is hardly a question to which Freudian theory is quite irrelevant. The theory of the unconscious has been advanced and the techniques of analysis developed by such distinguished disciples as his daughter Anna Freud, Melanie Klein (a specialist in the analysis of children), and many others in Europe and the United States. Theory and techniques have also been developed by many disciples and eclectics. Two of Freud's disciples who deviated from his theories—Alfred Adler and Carl Jung—have had very considerable influence.

Alfred Adler (1870–1937) distinguished his system from psychoanalysis by labeling it individual psychology. Before meeting Freud, he had made a special study of the biological phenomena of compensation for defective bodily organs. After his association with Freud he extended his principles to account for all forms of compensation for inferiority, the “inferiority complex.” In deviating from Freud, he assigned less importance to unconscious motivation and to sexuality.

Carl Gustav Jung (1875–1961) labeled his system analytical psychology. He differed from Freud in assigning a less important place to sexual motives and in his account of the unconscious. Jung regarded the libido as an undifferentiated “life force” which became differentiated into a number of instincts or drives. In his long life Jung developed a number of important but highly controversial theories. He elaborated the controversial and obscure concept of the collective unconscious and a theory of archetypal ideas (which has been confused by some with the Platonic concept of archetypes). Less controversial were the results of his experimental studies of word association and his suggestions regarding personality types. His wide-ranging speculations covered alchemy,

mythology, and the psychology of religion. Students of religion have found in Jung much of what they found absent or uncongenial in the writings of Freud. The opposition between Freudian and Jungian psychology has provided a modern parallel to the classical distinction between the Aristotelians and the Platonists.

*Horomic psychology.* In the Wundtian system as interpreted by Titchener the elements of mind were sensations and feelings. Conative experience had been eliminated by reductive analysis. Similarly, the concept of conative behavior had no place in Watsonian behaviorism. The concept of conation was not prominent in early Gestalt theory. Before 1950, however, the concept of conative or goal-directed behavior had been restored as a key concept in most systems of psychology. Tolman, the most sophisticated of the self-proclaimed behaviorists, established a new purposive behaviorism, and Lewin steered Gestalt psychology into the study of volitional processes. Throughout, Freudian theory is permeated by the facts of goal-directedness. The most thoroughgoing exponent of a conative psychology was William McDougall (1871–1938).

McDougall had a medical education but devoted himself to research in physiology, making several significant discoveries. An important early publication was his brief *Physiological Psychology* (1905), which contains the germs of his later theories. His most important publication was his *Introduction to Social Psychology* (1908). This title was unfortunate since the book contains the essentials of his general theory of motivation. Central to this theory was the thesis that there is a limited number of prime movers by whose conative force every train of thought and every bodily activity is initiated and sustained. These prime movers were first described as instincts, but the objections that were raised to his extreme deviation from the traditional biological conception of an instinct led McDougall to redescribe them as propensities.

In his detailed elaboration of these “propensities” McDougall developed an account of instinctive behavior originally suggested by William James. Prior to James instinct had been regarded as a biological mechanism producing rigid and stereotyped forms of behavior that were neither learned nor modified by experience. James drew attention to the cognitive emotional and impulsive components in instinctive action. McDougall developed this idea within the framework of the tripartite analysis of conscious experience that he had learned from Stout. Stressing the extent to which instinctive dispositions are modified both on the cognitive (receptive) side and on

the conative (responsive) side, he suggested that the primary instincts are to be defined by reference to the central or affective components, the “primary emotions.” He went on to describe the ways in which instinctive dispositions are modified and the ways in which they are organized into more complex motivating dispositions, the sentiments. A sentiment was conceived of as a system of instinctive disposition organized around an idea. Patriotism, for example, is a complex organization of instincts directed to promoting the welfare of a national group. McDougall’s account of the structure of human personality was similar to that first set out in the famous sermons of Bishop Butler on human nature (1726). With McDougall, as with Butler, the motivating forces in man are organized in a three-tiered hierarchy. At the base are the primary instincts or propensities. At the second level in Butler’s system were certain regulating and controlling principles, such as benevolence and cool self-love, and at the summit was the ultimate controlling principle, which was identified with conscience.

In McDougall’s system the basic instincts are organized into and controlled by the sentiments, which function in a similar way to Butler’s principles of benevolence and cool self-love. Thus, the parental sentiment is an organization of the maternal instinct together with other instincts, and in McDougall’s view it explains all disinterested altruism. The self-regarding sentiment is an organization of the instincts of self-assertion together with others that exercise a similar control over primitive aggressive instincts. It functions in McDougall’s theory in a way similar to Butler’s cool self-love and Freud’s reality principle. At the head of the hierarchy in McDougall’s system as the supreme controlling force is a master sentiment that is an elaborated form of the sentiment of self-regard.

Both Butler’s and McDougall’s accounts of the structure of human personality, of human motivation, and of the basis of volition or self-control have important similarities with, but also important differences from, Freud’s hierarchy of id, ego, and superego. Butler’s analysis had greater philosophical subtlety than McDougall’s, but McDougall’s was developed in much greater detail. The central theses were contained in *Social Psychology*. The details were further elaborated in his later works, such as the *Outline of Psychology* (1923) and the *Outline of Abnormal Psychology* (1926). McDougall was himself surprised, as well as gratified, by the outstanding success of his *Introduction to Social Psychology*. He was to be surprised and disappointed by the reception of what he intended to be his magnum opus, *Body and Mind: A History and*

*Defense of Animism* (1911). This contained a critical review of the traditional theories of the relations of body and mind in which he eventually decided in favor of interactionism. His general philosophy of nature, life, and mind was that of an orthodox dualist and interactionist. This was later developed into a Leibnizian monadology. The personality of man was conceived as a hierarchy of monads. Every monad is potentially a thinking, striving self, but each differs in degrees of development. At the head of the hierarchy is the supreme monad—the self, which is in command of, and directly or indirectly in communication with, all subordinate monads. The mode of communication was conceived to be telepathic.

McDougall was one of the last of the academic psychologists to attempt a comprehensive system covering all the facts of cognition, feeling, and conation as well as the facts of unconscious motivation. His theories, however, fell out of favor, though not entirely because of specific objections to them. They were outmoded by current trends in both psychology and philosophy. Nevertheless, he exercised a considerable influence on thought and research in motivation theory, not least upon those who differed from him, and he contributed to the reunification of psychology and the biological sciences, which had been separated since Aristotle's day. Indeed, it could be argued that McDougall, like Aristotle, saw that the concept of purpose was both logically irreducible to mechanistic concepts and fundamental for the explanation of human behavior. His mistake was to translate this eminently defensible conceptual doctrine into a genetic doctrine about the origins of behavior. The two do not necessarily go together, for the doctrine that human behavior cannot be explained without recourse to a concept like purpose does not entail the genetic doctrine that men must come into the world equipped with a myriad of built-in purposes.

REACTION AGAINST REACTIONS. The proliferation of schools continued into the 1930s. Carl Murchison's *Psychologies of 1925* was followed by his *Psychologies of 1930*, and at the time no end to such quinquennial volumes could be foreseen. Psychologists, however, began to tire of these battles among the schools, each of which was in revolt against the established order and at war with the others in revolt. There came a revolt against revolt, a reaction against reactions. Robert S. Woodworth (1869–1962), who had written the most influential critical commentary on the schools, *Contemporary Schools of Psychology* (1931), was a leading advocate of a middle-of-the-road psychology. Teaching and practicing psychologists tended to be eclectic; many leaned heavily on one or

another of the schools, and only a few remained uncommitted.

Schools were then replaced by “approaches,” a term that suggests convergence rather than divergence. Approaches, like viewpoints, are complementary. The new situation favored the emergence of groups of psychologists united in discipleship to a single dominating personality. These groups differed from the schools in that a school was created by several outstanding personalities who, though agreeing on certain basic theses, made individual contributions to the system of psychology defended by the school. There have always been groups of the simpler leader-and-disciples type. Before the age of the schools there were philosophical psychologists with their disciples—for example, Brentano and Alexius Meinong on the Continent, Ward and Stout in Great Britain, James in the United States. In the schools themselves there were subgroups composed of a man and his disciples—the Freudians, the Jungians, the Pavlovians, and so on. After the dissolution of the schools new personalities emerged, each with an individual approach or field of specialization; there were psychologists like Piaget at Geneva, Albert Michotte at Louvain, and Tolman and many others in the United States.

RELATION TO PHILOSOPHY. The history of psychology in the twentieth century is a story of the divorce and remarriage of psychology and philosophy. The trouble began when psychologists claimed the status of empirical scientists. At first the philosophers were the more aggressive, deriding the young science as a bogus discipline. The psychologists hit back and made contemptuous remarks about philosophical logic-chopping and armchair psychology. The arguments were charged with emotion, and neither side emerged with great credit. Slowly, some progress was made toward a diagnosis of the situation, a diagnosis that may well provide the basis for a happy reconciliation.

Psychology has always been, and may well always remain, a parasitic discipline. For twenty centuries it was just a branch of philosophy. To gain emancipation, it entered into willing bondage to the established natural sciences. Increasingly it has claimed to be, and has been increasingly accepted as, a biological science. Aristotle's psychology had a biological orientation, and theories of the temperaments have always had a physiological slant. Since Darwin psychologists have attempted to work down to the biological foundations of mental life, and biologists have extended their field of interest upward to include the more complex functions of organisms tradi-



tionally described as mental—perception, learning, problem solving. In the twentieth century psychologists and biologists found a common approach, frame of reference, and interest in such new special studies as ethology, cybernetics, and information theory and a common lack of interest or only a peripheral interest in problems of the philosophy of mind. There have, however, been other developments that have helped to resolve the conflicts between philosophers and psychologists and to clarify the lines of demarcation between work that can properly be done in an armchair and work that must be done on a laboratory stool, in a birdwatcher's blind, or behind a one-way screen.

The behaviorists, in their revolt against Titchener's introspectionism, had taken over quite uncritically Titchener's greatest error. Hegel had attempted to answer questions of empirical fact by a priori reasoning. Titchener made the opposite mistake, supposing that questions of philosophical analysis could be settled by observations made in a laboratory. His mistake is on record; he recalled that in 1888, when first reading James Mill's *Analysis of the Human Mind*, the conviction flashed upon him, "You can test all this for yourself." He thought he could test it by introspection. The *Analysis* of James Mill was an exercise in philosophical analysis that can be carried out in a soft armchair, perhaps more efficiently there than on a hard laboratory stool. The behaviorists also fell victim to the same error in confusing introspection and philosophical analysis, in failing to see that questions of analysis arise not only in regard to introspective reports but also in regard to behavioral concepts—stimulus, response, and behavior itself.

However, behaviorists and other biologically minded psychologists were little disposed to either philosophical speculation or philosophical analysis. They were content, like most biologists, to think of the world, regardless of consistency, *both* in terms of commonsense realism and in terms of the billiard-ball atomism of nineteenth-century physics, thereby following the physicists whenever they revised their theories. Those who had some interest in philosophy followed the prevailing trend in philosophy to some form of phenomenalism.

**Reduction of mental concepts.** There had been three centuries of philosophical thinking devoted to the elimination of superfluous psychological concepts. At first a mind was thought of as an immaterial substance that, like a material substance, persists through changing states. As a rod of iron passes through states of being hard and soft or black, red, and white in accordance with changes of temperature, so a mind passes through states of joy, sor-

row, and so on in accordance with the success and failure of its endeavors. Descartes had described all modes of consciousness as states of the soul, some of which appear to be states of external bodies, others of which appear to be states of the body in which the soul is embodied, and others that really are, as they appear to be, states of the soul itself. In his new way of ideas Locke redescribed experience in terms of the soul, self, or ego being presented with and attending to objects in the mind that chiefly represent things in the external world. Berkeley pointed out, cogently, that there is no way of comparing these representative ideas with the things they are supposed to represent. There were, he suggested, no reasons for, and there were reasons against, supposing that there are material things to be represented. Exit the material world. Then came Hume, who gave an important negative introspective report. He could not observe this soul, self, or ego to which presentations were said to be presented. Exit the soul.

For a long time attempts were made to defend what Titchener described as an act and content psychology—the doctrine that mind consists in mental contents and acts of willing and attending concerned with these contents (without, however, anyone to perform these acts). Late in the nineteenth century Brentano argued that these acts or attitudes are what is distinctive of mind. G. E. Moore based his refutation of idealism on this thesis by distinguishing in sensation the sensing, which alone is distinctively mental, from the sense datum sensed. But, like Hume, he made another negative introspective report—that the act is diaphanous, unintrospectable. Exit the act, the last claimant to mentality.

This reduction and elimination acquired a temporary finality in Bertrand Russell's neutral monism. Influenced by Moore, Ernst Mach, and William James, he proposed the overall theory that the stuff of which the universe is composed is neutral, not mental or physical. Organized in one way, it issues in the laws of physics; organized in another way, it results in the laws of psychology. Combining these, we have an account of nature. In this long reductive process man first had lost his soul, then his mind, then his consciousness, and finally even his body, which was reduced to a permanent possibility of neutralized sensations.

**Linguistic approach.** The finality of this form of phenomenalism was short-lived. The conception of philosophy as an inquiry into the ultimate nature of reality was supplanted by the idea that philosophy is the critical analysis of the concepts of science and of common sense. This was in turn replaced by the idea of philosophy as the

study of linguistic usages. Instead of asking what mind is, philosophers set out to disentangle the various uses of the word *mental*, and they became interested in the depth psychologists' uses of new words and of old words in new senses. Philosophers and psychologists began to find a new basis for collaboration. The philosophers clarified concepts; the psychologists attempted to verify by laboratory procedures the hypotheses stated in these concepts.

Not all issues between philosophers and psychologists have been resolved, but there has been notable progress toward a policy of coexistence, and here and there some progress toward cooperation has been made.

**See also** Animal Mind; Apperception; Behaviorism; Consciousness; Dreams; Emotion; Existential Psychoanalysis; Experience; Gestalt Theory; Guilt; Happiness; Humor; Images; Imagination; Intention; Intuition; Memory; Mind-Body Problem; Pain; Perception; Pleasure; Psychoanalytic Theories, Logical Status of; Religion, Psychological Explanations of; Sound; Thinking; Time, Consciousness of; Touch; Unconscious; Volition.

### **Bibliography**

- Boring, E. G. *A History of Experimental Psychology*. 2nd ed. New York: Appleton-Century-Crofts, 1950.
- Boring, E. G. *Sensation and Perception in the History of Experimental Psychology*. New York: Appleton-Century, 1942.
- Boring, E. G. et al. *History, Psychology, and Science: Selected Papers*. New York: Wiley, 1963.
- Flügel, J. C. *A Hundred Years of Psychology, 1833–1963*. London: Macmillan, 1933. Enl. ed., New York: Basic, 1964. The enlarged edition has an additional part, by Donald J. West, on the period from 1933 to 1963.
- Hamlyn, D. W. *Sensation and Perception*. London: Routledge and K. Paul, 1961.
- Hearnshaw, L. S. *Short History of British Psychology, 1840–1940*. New York: Barnes and Noble, 1964.
- Murphy, Gardner. *Historical Introduction to Modern Psychology*. Rev. ed. New York: Harcourt Brace, 1949.
- Peters, R. S., ed. *Brett's History of Psychology*. Rev. ed. London and New York, 1962. Consult pp. 769–772 for a bibliography of all major texts in the history of psychology. This work is an abridgment and updating of G. S. Brett's three-volume *History of Psychology* (London: Allen, 1912) and is the only complete history of psychology.

**R. S. Peters (1967)**

**C. A. Mace (1967)**

## PSYCHOLOGY [ADDENDUM]

In the 1950s and 1960s, scientific psychology underwent a major transformation. Behaviorist, Gestalt, and Freudian views were largely superseded by an approach called cognitive psychology, which treats the mind as a kind of information processor analogous to a computer. Cognitive psychology investigates the mental structures and processes that underlie perception, attention, learning, memory, language, inference, and problem solving. The field retains some behaviorist, Gestalt, and Freudian insights, but provides a coherent alternative that has been highly fruitful both experimentally and theoretically.

### THE COGNITIVE REVOLUTION

The roots of cognitive psychology lie partly in the limitations of previous theoretical approaches to psychology, particularly behaviorism. Behaviorism attempted to make psychology scientific by avoiding reference to hypothetical mental entities such as thoughts and concepts. It tried to restrict psychology to the use of observed stimuli to predict observed behavioral responses. Behaviorism was fueled in part by a positivist philosophy of science that failed to recognize that explanation in natural science abounds with hypothetical entities such as atoms and genes. By the 1950s it was becoming apparent that stimulus-response relations were inadequate to account for human verbal behavior and even for learning in rats.

The emergence of an alternative explanatory framework came from several sources. One was information theory, developed by Claude Shannon in the 1940s, which inspired psychologists such as George Miller to try to characterize the capacities of the human mind to process information. Miller's 1956 paper, "The Magical Number Seven Plus or Minus Two," reviewed evidence that minds are inherently limited in their ability to hold only a small amount of information but argued that this limitation is surmounted by representations that chunk pieces of information together. Cognitive psychology has largely abandoned the information-theoretic division of information into discrete bits, but the metaphor of information processing remains pervasive.

A second and ultimately more important source of cognitive psychology was the development in the late 1940s and 1950s of the idea of a computer program. Before the advent of computers, philosophers and psychologists who wanted to give a mechanistic account of mind were limited to relatively simple mechanisms such as clockworks and telephone switchboards. Computer

programs consist of representational structures such as numbers, words, and lists, along with algorithmic procedures that transform the structures to produce new ones. In their 1960 book *Plans and the Structure of Behavior*, George Miller, Eugene Galanter, and Karl Pribram compared the plans that control behavioral operations to computer programs. By 1956, Allan Newell and Herbert Simon had already developed a computer program that could simulate human performance on a high-level task, proving logic theorems. Most theories in cognitive psychology operate with the analogy that human thought applies mechanical processes to mental representations, just as computation applies algorithms to defined structures. In the strongest view, thinking is not just *like* computation, it *is* a kind of computation.

The third conceptual source for cognitive psychology was Noam Chomsky's new approach to linguistics, developed in the 1950s as an alternative to the behaviorist approaches of Zelig Harris that then dominated the field. Chomsky incisively criticized the explanatory adequacy of behaviorist accounts of language learning. He proposed an alternative that postulated mental structures such as an innate universal grammar that makes possible the efficient learning of any human language.

Ideas about information, computation, and mental grammars redirected the experimental research that occupies most psychologists much more than do theoretical matters. Investigation shifted from studies of animal behavior to experiments with human subjects concerning such mental operations as visual pattern recognition, memory, verbal learning, and speech perception. In 1967 Ulric Neisser published the new enterprise's first textbook, *Cognitive Psychology*, and the journal of the same title began three years later. Neisser focused on processes for visual and auditory cognition.

## TOPICS IN COGNITIVE PSYCHOLOGY

Later textbooks have addressed a broader range of topics, especially learning, memory, attention, perception, problem solving, language, representation, decision making, and deductive and inductive inference. Many experimental results have accumulated concerning these cognitive processes.

Research on perception has investigated how people recognize objects and other structures such as faces. Perceptual recognition involves both bottom-up processing from physical stimuli registered on sensory receptors such as the retina, and top-down processing influenced by high-level beliefs and concepts. Visual imagery has been a lively area of research, with Stephen Kosslyn and

others arguing that evidence supports the view that minds operate with visual as well as verbal representations. The study of attention considers the factors that lead people to focus on and shift their concentration to different aspects of their environment.

Memory researchers distinguish between long-term memory, which permanently stores representations of events and concepts, and working memory, which holds and manipulates information as people perform cognitive tasks. Human memory is very different from computer memory, which stores information exactly as presented to it. Consolidation of events and facts into long-term memory involves reconstruction and blending with previous experience. People are conscious of only part of the contents of working memory and are totally unaware of most of the cognitive processing that constitutes thought. Almost all thinking is unconscious, not because of Freudian repression mechanisms, but because people have little access to most of the operations of their brains.

Cognitive psychologists distinguish between episodic memory for particular events and semantic memory for conceptual relationships. Debate has raged concerning what concepts are, although most psychologists reject the traditional view, still found in philosophy, that concepts can be defined by necessary and sufficient conditions. Alternative theories of concepts maintain variously that they consist of prototypes that specify typical but not universal features, sets of exemplars of objects, general knowledge about things, or patterns of activation in neural networks. These theories have inspired experimental investigations of how concepts are learned.

Cognitive psychologists who study language perform experiments concerning how people comprehend and produce utterances. In the 1960s and 1970s, research in psycholinguistics was heavily influenced by Chomsky's theory of transformational grammar, but later research shifted away from emphasis on the syntactic structure of language to concern with its meaning and communicative functions. There has also been investigation of high-level linguistic processes such as reading and understanding discourse.

Like the Gestalt psychologists, cognitive psychologists have been interested in problem solving and creativity. They have constructed detailed computational models of how people solve different kinds of problems using general rules and/or analogies with previous problem solutions. Theories of expertise have been developed based on different accounts of how different mental

structures constitute the knowledge possessed by people with substantial experience in a particular domain.

Psychological investigations of inference have been at odds with normative models of reasoning popular in philosophy. In contrast to the view that deductive inference involves the application of formal syntactic rules such as those found in propositional logic and predicate calculus, much psychological research has supported the view that the human mind makes inferences in a way that does not sharply distinguish syntax from semantics. People make deductive inferences based on the content of representations, not just their form. Similarly, people do not make decisions using the formalism of expected utility theory, which sharply distinguishes utilities from probabilities. Instead, they use a variety of cognitive and emotional heuristics to evaluate and choose different options. It is controversial whether the deviation of people from normative models of inference shows that they are irrational, or whether there is a need for more psychologically realistic models of rationality.

## THEORETICAL AND EXPERIMENTAL DEVELOPMENTS

In the 1970s, cognitive psychology became part of the interdisciplinary field of cognitive science, which also includes philosophy, linguistics, anthropology, neuroscience, and the branch of computer science called artificial intelligence. Computational ideas continued to be at the core of psychological theory. Some theorists such as Allen Newell and Herbert Simon, and later John Anderson, maintained that psychological phenomena are best explained by postulating that the mind uses rules of the form, IF such and such, THEN such and such. These rules are operated by procedures that search through a large number of possible sequences of operations to provide solutions to problems. Rule-based models have been used to provide detailed explanations of problem solving, skill acquisition, and language production and acquisition.

Other psychologists have emphasized the role that structures such as concepts and schemas play in cognition. From this perspective, problem solving is not so much a sequential search through a large range of possible moves, as the application of patterns that enable people to comprehend and respond to situations. Thought is viewed as a kind of pattern matching rather than as a search through a space of operators defined by rules. Computational models of visual processing have inspired new theories and experiments concerning the nature of visual imagery.

During the 1980s, there was an influx of theoretical ideas based on computational models of artificial neural networks. This approach is called *connectionism* because it emphasizes the connections (links) between simple neuron-like processors. Inferences of the sort performed by rules and concepts are supposed to emerge from the interactions of many highly connected units that take in activation from many other units and pass it on to many others. Learning consists in adjusting the strengths of the links between the units. Connectionist models have been applied to many psychological phenomena, including learning from examples and high-level problem solving of the sort discussed by Gestalt psychologists.

From psychology journals such as *Cognitive Psychology*, it is evident that most psychological research is experimental rather than theoretical. Experimenters use a variety of measures such as error rates and reaction times to detect characteristics of human thinking. For example, investigations of eyewitness memory and testimony show that it is sometimes inaccurate, with errors being more likely if the witness has been distracted, if the misinformation is plausible, if there is social pressure, and if eyewitnesses have been given surreptitious positive feedback. Studies of deductive and inductive inference look at the difference between examples where people reason well and examples where they tend to make mistakes. Another experimental measure is reaction time, which compares how fast people are to respond to different stimuli. For example, people are quicker to respond to an item if it is preceded by a similar item that primes it.

In the 1990s, cognitive psychology began to draw much closer to neuroscience. Cognitive psychologists have always assumed that mental operations are carried out by the brain, but for decades they lacked experimental techniques to relate human behavior to brain processes. Instruments are now available for imaging what is happening in the brain while people perform cognitive tasks. The most commonly used are PET (positron emission tomography) and fMRI (functional magnetic resonance imaging) scans. Both tools can detect the increase in blood flow to regions of the brain that become active when it is performing a particular task. For example, brain scans can determine what parts of the brain have increased activity when people are asked to rotate mental images. Many cognitive psychologists have turned to performing experiments in which people perform mental tasks while their brains are being scanned.

These experiments have furnished data that are used to suggest and evaluate theories about what brain processes are involved in various cognitive tasks. Increas-

ingly, cognitive theorizing is more brain-oriented, invoking processes that occur within particular brain regions and involve the interactions of multiple brain regions. In contrast to the connectionist models of the 1980s, computational models based on cognitive neuroscience employ much more neurologically realistic ideas about the structure of brain networks and the operation of individual neurons. Whereas cognitive psychology originally ignored the role of motivation and emotion in human thinking, cognitive neuroscience has inspired new models that integrate cognition and emotion in accounts of human decision making. Even the topic of consciousness, assailed by the behaviorists as inherently unscientific, is now being investigated by means of psychological and neurological experiments along with neurocomputational theories.

Cognitive psychologists disagree about the extent to which mental structures and processes are innate. Psycholinguists and evolutionary psychologists such as Stephen Pinker argue that natural selection has furnished the mind with many special purposed inference mechanisms such as a language acquisition device. The alternative view is not that the mind is a blank slate with no innate machinery, but rather that what all humans inherit is a highly flexible learning mechanism that makes possible adaptation to many situations. Although proponents of innateness are more likely to advocate rule-based rather than connectionist theories, there is no inconsistency in maintaining that most rules are learned and that some connections are innate.

## OTHER AREAS OF PSYCHOLOGICAL RESEARCH

Cognitive psychology is only one area of psychological research, but it has had a major impact on other areas such as developmental, social, organizational, educational, and clinical psychology. Developmental psychologists study the origins and growth of children's knowledge of language and the world. The ideas of Jean Piaget dominated developmental psychology for decades, but they have been reassessed and revised by means of theories and experiments suggested by cognitive psychology. Developmental psychologists have also constructed new experimental techniques, such as the measurement of infants' attention times used to indicate what kinds of objects and situations are unfamiliar to them.

Social psychology, which concerns how people interact with each other, has become dominated by the field of social cognition, which examines people's cognitive processes. Social cognition investigates how people make

sense of each other using concepts, stereotypes, rules, hypotheses, memories, emotions, personality traits, and other mental representations. Social psychology also looks at how cognition can vary across different cultures. Cognitive psychology has had an equally major impact on organizational psychology, which studies such topics as management and industrial development. In particular, work on organizations has been influenced by cognitive theories of decision making and learning. Similarly, cognitive theories of learning such as ones concerned with the acquisition of rules and concepts have had a substantial impact on educational psychology.

Clinical psychology, another area with practical applications, has also been transformed by the cognitive revolution. This area was once dominated by Freudian theories, which generated little success in either experimental or clinical settings. Cognitive therapy is one of the few kinds of psychotherapy that have been shown in controlled experiments to benefit people with emotional disorders such as depression. Unlike psychoanalysis, it does not require detailed discussion of a patient's childhood, but instead concentrates on helping the patient to replace unrealistic beliefs and goals in order to produce more positive appraisals of themselves and their situations. The medical field of psychiatry has also abandoned Freudian theories in favor of neurochemical explanations and treatments of mental illnesses such as depression, mania, dementia, and schizophrenia.

## CONTROVERSIES

Although psychology has accumulated an impressive body of experimental findings and theoretical explanations, there remain topics of controversy that indicate directions of future work and generate interesting philosophical issues. These topics include: nature versus nurture, culture versus universality, rules versus connections, images versus propositions, mental logic versus mental models, heuristics and biases versus the adaptive toolbox, and embodiment versus computation.

**NATURE VERSUS NURTURE.** Psychologists disagree about the extent to which the behavior of humans is the result of innate, genetically transmitted neural structures (nature), compared with the extent to which it is the result of learning from physical and social environments (nurture). The nature side is currently emphasized by evolutionary psychologists such as Stephen Pinker and Leda Cosmides, who argue that many specific mental abilities have developed as the result of evolution by natural selection. They have proposed that the brain con-

tains evolved modules that are specialized in function, such as a face recognition system, a language acquisition device, navigation specializations, and a routine for detecting cheaters in social situations. The alternative view is not that there are no genetically inherited structures in the brain, but rather that the main innate endowment is a flexible learning ability that enables people to adapt to a wide range of environments. Proponents of this view include Jeffrey Elman, Steven Quartz, and Terrence Sejnowski. Psychologists also debate the extent to which general intelligence is inherited, and even whether there is such a thing in contrast to specific kinds of intelligence such as verbal, social, and physical abilities. Resolution of these debates will require further research on the structure and development of the brain.

**UNIVERSALITY VERSUS CULTURE.** Psychologists have tended to assume that the mental processes they investigate are universal, operating in all human minds. In contrast, anthropologists have tended to emphasize the diversity of different cultures with respect to beliefs and practices. An increasing number of psychologists have been using experimental methods to investigate the impact of culture on cognition, motivation, and emotion. Richard Nisbett and others have explored the impact of cultural differences on aggressive behavior and even on general styles of thinking. Compared to Westerners, East Asians are more likely to notice environments and relations rather than objects, to see change rather than stability, and to explain other people's behavior in terms of situations and relationships rather than personality traits. This issue is not the same as the nature versus nurture issue because there is no evidence of any relevant biological differences between Westerners and East Asians; cognitive differences therefore reflect culture rather than genetics. Further cross-cultural work in cognitive and social psychology will provide more information on the extent to which cognitive processes are universal.

**RULES VERSUS CONNECTIONS.** Many psychologists follow Chomsky in supposing that the acquisition and use of language depends on the possession of rules. For example, the standard way to form the past tense in English is to add "ed," as in "Sheila argued with Tom." But there are also many exceptions to this rule, such as "threw" and "went." James McClelland and other connectionists have argued that language use does not require rule acquisition and can be understood as the result of learning mechanisms that modify the links in neural networks. In contrast, Stephen Pinker and others argue that rules are necessary to explain patterns of linguistic behavior.

**IMAGES VERSUS PROPOSITIONS.** Philosophers such as Jerry Fodor have interpreted the computational view of mind to imply the existence of a language of thought that uses a representational scheme akin to verbal propositions. In contrast, most psychologists maintain that human thinking involves more than one kind of code, in particular visual images in addition to verbal propositions. Stephen Kosslyn and other have argued that a combination of psychological and neurological evidence, along with computer simulations, support the hypothesis that minds use visual as well as verbal representations. For example, brain scanning experiments show that when people perform imaging tasks, they use parts of the brain involved in visual processing. However, Zenon Pylyshyn continues to maintain that the evidence does not support the hypothesis that visual imagery is computationally different from verbal inference.

**MENTAL LOGIC VERSUS MENTAL MODELS.** How do people perform inferences such as the following? All humans think; anything that thinks has a brain; so all humans have brains. Philosophers since Aristotle have used formal logic to identify valid deductive inferences that accord with rules of inference such as *modus ponens*. Some psychologists such as Lance Rips similarly argue that human inference uses a kind of mental logic based on abstract rules. In contrast, Philip Johnson-Laird and others have presented experimental evidence that human inference does not distinguish form and content, but rather works with concrete instantiations that he calls *mental models*. The mental model approach has been applied to syllogistic, propositional, probabilistic, and causal inferences.

**HEURISTICS AND BIASES VERSUS THE ADAPTIVE TOOLBOX.** Many philosophers since Aristotle have assumed that humans are inherently rational, but many psychologists have investigated common tendencies to make inferential errors in inductive reasoning. Daniel Kahneman and Amos Tversky launched a fertile research program that showed that people often have difficulty making inferences that accord with normative models based on mathematical theories of probability and utility. They proposed instead that people operate with simple mental heuristics that bias them into making inferential errors. For example, people might think there are more words that start with "R" than end with "R," because it is easier to think of examples of the former rather the latter. Gerd Gigerenzer agrees with Kahneman and Tversky that human rationality is bounded by cognitive limitations but argues that biological evolution has provided people

with heuristics that are fast, frugal, and quite effective. For example, he thinks that people have heuristics for reasoning about frequencies that can enable them to avoid many of the reasoning errors identified by Kahneman and Tversky.

**EMBODIMENT VERSUS COMPUTATION.** Since the cognitive revolution in the 1960s, most psychologists have adopted the view of mind as an information processor analogous to a computer. An alternative view, based on the work of James Gibson and others, is that much of human behavior can be understood in terms of responses by minds to properties of their physical and social environments. On this view, human thought depends heavily on the kinds of bodies that people have, with sensory and kinesthetic abilities that shape our perceptions and thoughts. Extreme versions of this view deny that computation is part of human thinking at all, but more moderate views emphasize special kinds of computations performed by brains that store information using visual and other formats that are tied to the sensory apparatus of human bodies. Then the embodied nature of much of human thinking is a useful supplement to the computational theory of mind, not an alternative to it.

### RELATIONS TO PHILOSOPHY

Developments in cognitive psychology have been important to philosophy for two reasons. First, philosophy of mind has responded to changes in the nature of psychological theories and explanations, with implications for metaphysical issues concerning the relation of mind and body. Second, a major strain of epistemology has become naturalistic, viewing theories of knowledge as continuous with cognitive psychology. Philosophical naturalism has also had an impact on ethics.

**PHILOSOPHY OF MIND AND PSYCHOLOGY.** The common-sense theory of mind is dualism, according to which people consist of two substances, matter and soul. In contrast, materialism maintains that humans, like the rest of the universe, consist only of matter and energy. Materialists from Epicurus to Lucretius to Thomas Hobbes have faced the daunting project of explaining how mental phenomena such as thinking and consciousness arise from brain activity. The development of cognitive psychology, with its view of thinking as computation performed by the brain, has greatly contributed to philosophical theories of mind.

In the 1960s and 1970s, philosophers such as Hilary Putnam, Jerry Fodor, and Daniel Dennett discussed the

implications of the new ideas about computation and cognition. In the 1950s, materialist philosophers such as J. J. Smart had advocated an *identity* theory, according to which mental processes are brain processes. But the advent of artificial intelligence and computational theories of mind suggested an alternative view that mental processes are independent of any particular physical realization. This view is called *functionalism* because it understands mental states as functional states that are related to each other computationally. Functionalism is still a version of materialism, because it assumes that mental states have a basis in brains, computers, or some other form of matter and energy. But it differs from the identity theory in not equating mental states with any particular kind of physical state.

Functionalism gained much plausibility from the rise of computational views of mind, but it has been challenged by developments in cognitive neuroscience. Cognitive psychology has moved away from abstract computational theories toward theories embedded in particular accounts of the structure and processes in the human brain. It has therefore become less plausible that mental states are functional states rather than specific brain states. Moreover, progress in the field of artificial intelligence has not been as great as its originators had hoped. There have been some impressive industrial applications, but the prospect of a general-purpose machine intelligence comparable to humans remains distant. In contrast, understanding of how the brain uses the biochemical properties of neurons organized into connected functional areas has expanded rapidly. Hence developments in scientific psychology have lent support to identity over functionalist theories of mind. Some philosophers such as David Chalmers have argued that problems in understanding consciousness require a form of dualism, but many psychologists and neuroscientists remain optimistic that even consciousness will yield to scientific explanation.

What is the nature of the theories, explanations, and experimental results that cognitive psychologists offer? Philosophy of science has often adopted the model, derived from physics, that a theory consists of universal laws that deductively explain universal generalizations from observation. But the results of psychological experiments are usually statistically significant effects or tendencies, not laws. Moreover, psychological theories are rarely stated as universal laws, and the relationship between theories and what they explain is rarely deductive. Instead, a theory in cognitive psychology is a description of a computational or neurological mechanism,

where a mechanism is a system of parts whose properties and relations produce regular changes in other parts. For example, a bicycle is a mechanism consisting of parts such as pedals, wheels, a crank, and a chain that interact with each other. In purely computational theories such as those based on rules, the parts are mental representations, and the changes are brought about by computations on the representations. In neurological theories, the parts are neurons organized into brain areas and the changes are brought about by biochemical processes. Ideally, theories can be both neurological and computational when they show how groups of neurons implement mental representations and how biochemical processes implement computational operations.

**PHILOSOPHICAL NATURALISM.** Naturalism is the view that philosophical problems are continuous with those in science and amenable to treatment by the same kinds of methods used by scientists, as opposed to a priori theorizing about necessary truths and nonnatural entities such as gods and souls. In the late twentieth century, naturalized epistemology was revived through the work of philosophers such as Willard Van Orman Quine and Alvin Goldman. Quine's naturalism was limited by his adherence to behaviorist psychology, but later work has made full use of the expanding resources of cognitive psychology. Goldman has shown how to link philosophical questions about the origins and justification of knowledge with psychological research on perception, memory, and inference. This kind of naturalism does not use psychology to replace or reduce philosophy, which remains concerned with normative issues about justification that are not studied by psychologists. But naturalism applies epistemic appraisal to psychological processes that operate in human brains. Epistemic naturalism has also influenced the philosophy of science, with the view that scientific theories are mental structures rather than logical entities.

Moral naturalism has also been revived by philosophers who argue that ethics needs to pay close attention to cognitive psychology and neuroscience. According to moral naturalists such as Owen Flanagan, construction of moral theories and projection of moral ideals needs to ensure that the character, decision making, and behavior prescribed are possible for human beings. Understanding and evaluation of moral judgments is improved by appreciating how they arise from cognitive and neural processes such as concept application and empathy, which requires integration of cognitive operations with emotional processes. As with naturalized epistemology, moral naturalism does not purport to reduce ethics to

psychology, but rather to develop a richer account of moral justification consistent with rapidly increasing scientific knowledge about the cognitive and neurological sources of human action and judgment. Similarly, some philosophers have become involved in controversies concerning human rationality that arise from the debates about mental logic and heuristics and biases. According to the naturalistic perspective, epistemology, ethics, and metaphysics should continue to evolve hand in hand with further developments in cognitive psychology and neuroscience.

**See also** Anderson, John; Artificial Intelligence; Aristotle; Behaviorism; Chomsky, Noam; Cognitive Science; Computationalism; Dennett, Daniel Clement; Dualism in the Philosophy of Mind; Epicurus; Fodor, Jerry A.; Functionalism; Gestalt Theory; Goldman, Alvin; Hobbes, Thomas; Lucretius; Materialism; Memory; Naturalized Epistemology; Neuroscience; Piaget, Jean; Putnam, Hilary; Quine, Willard Van Orman; Smart, John Jamieson Carswell.

### **Bibliography**

- Anderson, John R. *Cognitive Psychology and Its Implications*. 5th ed. New York: Worth, 2000.
- Anderson, John R., with Francis S. Bellezza, et al. *Rules of the Mind*. Hillsdale, NJ: Erlbaum, 1993.
- Baars, Bernard. *The Cognitive Revolution in Psychology*. New York: Guilford, 1986.
- Bechtel, William, and Adele A. Abrahamsen. *Connectionism and the Mind: Parallel Processing, Dynamics, and Evolution in Networks*. 2nd ed. Malden, MA: Blackwell, 2002.
- Elman, Jeffrey L., Elizabeth A. Bates, Mark H. Johnson, Annette Karmiloff-Smith, Domenico Parisi, and Kim Plunkett. *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press, 1996.
- Gardner, Howard. *The Mind's New Science: A History of the Cognitive Revolution*. New York: Basic Books, 1985.
- Gigerenzer, Gerd. *Adaptive Thinking: Rationality in the Real World*. New York: Oxford University Press, 2000.
- Gilovich, Thomas, Dale Griffin, and Daniel Kahneman, eds. *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge, U.K.: Cambridge University Press, 2002.
- Goldman, Alvin I. *Epistemology and Cognition*. Cambridge, MA: Harvard University Press, 1986.
- Johnson-Laird, P. N. and Ruth M. Byrne. *Deduction*. Hillsdale, NJ: Erlbaum, 1991.
- Kosslyn, Stephen M. *Image and Brain: The Resolution of the Imagery Debate*. Cambridge, MA: MIT Press, 1994.
- Kunda, Ziva. *Social Cognition: Making Sense of People*. Cambridge, MA: MIT Press, 1999.
- Matlin, Margaret W. *Cognition*. 6th ed. Hoboken, NJ: Wiley, 2005.
- May, Larry, Marilyn Friedman, and Andy Clark, eds. *Mind and Morals: Essays on Cognitive Science and Ethics*. Cambridge, MA: MIT Press, 1996.



- Miller, G. A. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." *Psychological Review* 63 (1956): 81–97.
- Miller, G. A., E. Galanter, and K. Pribram. *Plans and the Structure of Behavior*. New York: Holt, Rinehart, and Winston, 1960.
- Murphy, Gregory L. *The Big Book of Concepts*. Cambridge, MA: MIT Press, 2002.
- Neisser, U. *Cognitive Psychology*. New York: Appleton-Century-Crofts, 1967.
- Nisbett, Richard E. *The Geography of Thought: How Asians and Westerners Think Differently ... and Why*. New York: Free Press, 2003.
- O'Reilly, Randall C., and Yuko Munakata. *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. Cambridge, MA: MIT Press, 2000.
- Pinker, Steven. *The Blank Slate: The Modern Denial of Human Nature*. New York: Viking, 2002.
- Quartz, Steven R., and Terrence J. Sejnowski. *Liars, Lovers, and Heroes: What the New Brain Science Reveals about How We Become Who We Are*. New York: Morrow, 2002.
- Rips, Lance J. *The Psychology of Proof: Deductive Reasoning in Human Thinking*. Cambridge, MA: MIT Press, 1994.
- Thagard, Paul. *Mind: Introduction to Cognitive Science*. 2nd ed. Cambridge, MA: MIT Press, 2005.
- Wilson, Robert A., and Frank C. Keil, eds. *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, MA: MIT Press, 1999.

**Paul Thagard (2005)**

## PSYCHOLOGY, GESTALT

See *Gestalt Theory*

## PUFENDORF, SAMUEL VON (1632–1694)

Samuel von Pufendorf, the German political and legal philosopher and historian, was born in Dorfchemnitz, in Meissen, Saxony, the son of a poor Lutheran pastor. A scholarship enabled Pufendorf to attend the famous Prince's School at Grimma. From 1650 to 1656 he attended lectures on Lutheran theology and Aristotelian philosophy at Leipzig. Somewhat later he studied contemporary philosophy at Jena, where he also read newly published books on mathematics and discovered the works of Hugo Grotius and Thomas Hobbes. At Jena he came in contact with Erhard Weigel, a former teacher of Gottfried Wilhelm Leibniz, whose strange but original method of teaching ethics "mathematically" made a lasting impression upon Pufendorf. To Weigel, Pufendorf owed the inspiration for his first work on the general

principles of law, *Elementorum Jurisprudentiae Universalis*. In 1658 Pufendorf became a tutor in the house of the Swedish ambassador to Denmark. When war broke out between Sweden and Denmark, Pufendorf was imprisoned for eight months, and it was during this imprisonment that he composed the booklet inspired by Weigel. Upon his release Pufendorf migrated in 1659 to the Netherlands, where the work was published in 1660.

On the recommendation of Grotius's elder son, Pufendorf was offered the chair of natural and international law at Heidelberg, the first such chair at a German university. He was soon appointed also as instructor of the heir to the crown of the Palatinate, and thus he began to mix with the electoral court, where he avidly studied the burning contemporary political problems. Out of this study came a pseudonymous work on the condition of the Holy Roman Empire, *De Statu Imperii Germanici* (1667), a work later famous for its statement that the constitution of the Empire resembles a monster, being neither a monarchy nor an aristocracy nor a democracy.

After his appointment as professor of natural law at the University of Lund in Sweden, Pufendorf wrote his fundamental work on national and international law, *De Jure Naturae et Gentium* (1672). The eight volumes of this compendium, which contains a veritable encyclopedia of the social sciences, are rather difficult reading. Pufendorf therefore produced an abstract of this work, titled *De Officio Hominis et Civis* (1673), which was soon translated into English, French, and German and thus found many readers abroad. By 1684 a Swiss Calvinist theologian was lecturing on the *De Officio Hominis* at Lausanne, but Lutheran theologians in both Sweden and Germany criticized Pufendorf's ideas vehemently. The king of Sweden himself had to protect his professor of law and induce the authorities of the university to defend Pufendorf against the charge of heresy. Pufendorf replied bitterly to the charge, and a long paper war ensued. Finally, Pufendorf published a "sanguinary" (his own description) polemical treatise titled *Eris Scandica* (Frankfurt, 1686), containing all his essays and letters relating to the controversy.

In 1677 Pufendorf was appointed by the king as court historian in Stockholm, where he spent ten years working on his extensive, thirty-three-volume history of Sweden, a work of no importance today except as an example of careful work and precise reporting. His shorter *Einleitung zu der Historie der vornehmster Reiche und Staaten* (2 vols., Frankfurt, 1682–1685) is more highly esteemed.

From 1688 until his death shortly after having been knighted by his former sovereign, the king of Sweden, Pufendorf lived in Berlin, where he had been called as court historian by the elector of Brandenburg.

A noted representative of the Baroque era, Pufendorf was a man of great self-confidence and stolid self-reliance. He had unshakable faith in the power of scientific reason and wished to establish it in the fields of jurisprudence and politics. He believed in the certainty of mathematics and rejoiced in the reunion of philosophy and mathematics then taking place. Although he wished to treat the problematic questions of jurisprudence and politics “mathematically,” he was a true empiricist who sought to introduce a “scientific” method into the study of history. He was therefore eager to undertake the thoroughly planned research into public archives that resulted in his history of Sweden.

Pufendorf thus united the two major trends of his age, Baconian empiricism and Cartesian logicism. One of the last polyhistorians, he united in his work all the methods of historical, sociological, and juridical thinking. A political figure rather than simply a lawyer, Pufendorf profoundly criticized the constitution of the Holy Roman Empire and its political conception. He argued for the founding of a European federation of sovereign states. He did not defend national or regional absolutism, however popular they were at the time; instead, he tried to unite the Hobbesian doctrine that the state should be governed by the rule of law and based on natural law in the empirical sense of the term (the war of all against all, *status necessitatis*) with the Grotian doctrine that the rule of international law should be based on natural law in an emotional sense (an inclination for society, *ordo amoris*). On this account Pufendorf has often been called a predecessor of eighteenth-century rationalism. Such a view is supported by his letter to his younger friend Christian Thomasius, in which he claimed that he “never had boldness enough to draw the utmost conclusions” from his philosophical rationalism and voluntarism.

Despite Leibniz’s opinion that Pufendorf was “a man of no great judgment,” his legal thought was of considerable importance and great philosophical interest. He was undoubtedly one of the most outstanding social philosophers on the European continent in the seventeenth century. It may be an exaggeration to call Pufendorf the first “philosopher of culture” (*Kulturphilosoph*) in Germany, but he was the first to grasp the fundamental concept of the sociological theory of law and politics. He saw the social realities of human life as a whole. His structural distinction between physical facts and moral institutions

inspired a new way of studying social facts in their independence and uniqueness. Following Weigel, Pufendorf distinguished four elements of social being: personality (*persona*), rank or profession (*status*), quality, and quantity. Every pattern of social order should be examined on the basis of these fundamental structures; for example, a state may be described in terms of its sovereignty, type of government, power, and population.

These elements, the ontological foundations of every community, have simultaneously to be interpreted as fundamental ethical principles of social life. Pufendorf designated three patterns of well-formed communities: humanity, ordered by the law of reason; Christianity, ordered by the law of God; and citizenship, ordered by the law of the state. Natural law, including religious and rational principles, therefore limits both civic and moral duties. Philosophy of law comprises both sociology and political science on the one hand, and jurisprudence and ethics on the other. This new discipline, which Pufendorf called simply natural law, was intended to unite all the tasks of interpreting social order and to combine the scholastic methods of the sixteenth-century Spanish thinkers with the newer ideas of Grotius and Hobbes.

In apparent contradiction to these sources of his thought on social order was Pufendorf’s strong belief in reason of state (*ratio status*). Although he often emphasized the self-determination and self-sufficiency of the state, he did not mean by this a totalitarian absolutism. And although he proclaimed the independence of political power against every ecclesiastical claim, he never taught the modern ideology of unlimited government, and his views were therefore not contradictory to the rule of law. What Pufendorf said about the relation of church and state must be interpreted dialectically. He conceded neither decisive authority to reason of state nor the right of moral constraint to the church.

Pufendorf may be called the initiator of the seventeenth-century movement of “scientific” natural law in Germany. By introducing the ideas of Grotius and Hobbes into German thought he made their ideas really effective for the first time. He liberated the natural-law theory from the domination of scholasticism and humanism. In so doing he built up an independent political science that always took into account contemporary history and reason of state. A clever and levelheaded politician, he predicted the decline of the Hapsburg monarchy after the Treaty of Westphalia. In criticizing the “monstrous” constitution of the empire he sought to advance a European commonwealth based on the natural and rational principles of international law. As a histo-

rian, Pufendorf introduced the empirical study of archives and gave an effective example of a new method of historical insight, and he may be regarded as an important predecessor of nineteenth-century historicism.

**See also** Aristotelianism; Cartesianism; Empiricism; Grotius, Hugo; Historicism; Hobbes, Thomas; Humanism; Leibniz, Gottfried Wilhelm; Natural Law; Rationalism; Scientific Method; Sovereignty; Thomasius, Christian; Voluntarism.

## Bibliography

### WORKS BY PUFENDORF

- Elementorum Jurisprudentiae Universalis Libri Duo*. The Hague, 1660.
- De Statu Imperii Germanici ad Laelium Fratrem, Dominum Trezolani, Liber Unus*. Geneva, 1667.
- De Jure Naturae et Gentium Libri Octo*. Lund, Sweden, 1672. Translated by Basil Kennet as *Of the Law of Nature and Nations*. London, 1710.
- De Officio Homini et Civis Prout Ipsi Praescribuntur Lege Naturali. Libri Duo*. Lund, Sweden, 1673. Reprinted with translation by F. G. Moore. 2 vols. Oxford, 1927. Translated by Andrew Tooke et al. as *The Whole Duty of Man according to the Law of Nature*, edited by Ian Hunter and David Saunders. Indianapolis: Liberty Fund, 2003.
- De Habitu Religionis Christianae ad Vitam Civilem*. Bremen, 1687. Translated as *Of the Relation between Church and State*. London, 1719.
- Of the Nature and Qualification of Religion in Reference to Civil Society*. Translated by J. Crull; edited by Simone Zurbuchen. Indianapolis: Liberty Fund, 2002.

### WORKS ON PUFENDORF

- Carr, Craig L., and Michael J. Seidler. "Pufendorf, Sociality and the Modern State." *History of Political Thought* 17 (3) (1996): 354–378.
- Dufour, Alfred. "Pufendorf." In *The Cambridge History of Political Thought, 1450–1700*, edited by J. H. Burns and M. Goldie. Cambridge, U.K.: Cambridge University Press, 1991.
- Hunter, Ian. *Rival Enlightenments: Civil and Metaphysical Philosophy in Early Modern Germany*. Cambridge, U.K.: Cambridge University Press, 2001.
- Krieger, Leonard. *The Politics of Discretion: Pufendorf and the Acceptance of Natural Law*. Chicago: University of Chicago Press, 1965.
- Scheuner, Ulrich. "Samuel Freiherr von Pufendorf." In *Die Grossen Deutschen*, Vol. 5, 126ff. Berlin, 1957.
- Schneewind, J. B. "Pufendorf's Place in the History of Ethics." *Synthese* 72 (1987): 123–155.
- Welzel, Hans. *Die Naturrechtslehre Samuel Pufendorfs*. Berlin, 1958.
- Wolf, Erik. *Grosse Rechtsdenker der deutschen Geistesgeschichte*, 4th ed., 311–366. Tübingen, 1963. Bibliography on pp. 367–370.

**Erik Wolf (1967)**

*Bibliography updated by Philip Reed (2005)*

## PUNISHMENT

The word *punishment* is used in varying contexts. The punishment meted out by the state to a criminal or by a parent to his children is not the same as the punishment boxers give or receive. The latter, however, is punishment only in a metaphorical sense, for it lacks several of the features necessary to a standard case of punishment. Characteristically, punishment is unpleasant. It is inflicted on an offender because of an offense he has committed; it is deliberately imposed, not just the natural consequence of a person's action (like a hangover), and the unpleasantness is essential to it, not an accidental accompaniment to some other treatment (like the pain of the dentist's drill). It is imposed by an agent authorized by the system of rules against which an offense has been committed; a lynching is not a standard case of punishment. Philosophers who have written on punishment have usually had in mind punishment in the standard sense rather than in any extended or metaphorical sense.

The philosopher's interest in punishment is mainly connected with questions of justification. It is, *prima facie*, wrong to deliberately inflict suffering or deprivation on another person, yet punishment consists in doing precisely this. What conditions, the philosopher asks, would justify it? Or, more generally, what kind of consideration would count toward a justification? For instance, if a person had already committed a crime, that would clearly be relevant to the question of whether he ought to be punished (although it might not be conclusive). What if he were only expected to commit a crime in the future? Or, again, is it relevant to the question of whether this man should be punished to say that punishing him would deter others? And assuming that criminals ought to be punished, how should we set about deciding appropriate penalties?

It is not, of course, the business of the moral or social philosopher to provide a justification for any particular act or system of punishment or even of the institution of punishment in general. Philosophers are not necessarily apologists for their society and age. They are interested in the procedures and modes of argument that we are committed to by our fundamental conceptions of morality and in criteria of criticism and justification rather than in inquiries into whether actual institutions satisfy them.

Philosophers, it is true, have not always made this distinction; they have often worked on the understanding that a philosophical argument could be seriously shaken by showing that it leads to conclusions inconsistent with some widely approved institution or moral rule. More-

over, for many philosophers, if such a rule or institution seemed to imply a principle inconsistent with other moral principles accepted by the society, there must necessarily be some broader principle, which a philosopher could discover and by which the conflict could be resolved. Applied to the case of punishment, this would mean that a philosopher must reconcile the apparently conflicting principles that wrongdoers should be punished and that it is wrong to deliberately make another person suffer. But this is surely a misconception of the nature of philosophy. There is no point, after all, in asking whether and how punishment can be justified if one assumed in advance that it can. For justification a number of contingent facts are required that the philosopher as such is not qualified to provide. His task is to analyze what is being asked for and so to point out what kinds of facts and arguments are admissible to the discussion.

### JUSTIFICATION OF PUNISHMENT

The question of justification arises at two levels. One can take for granted the principle that wrongdoers should be punished and ask whether a particular case of punishment was justified. At this level the philosopher is concerned with the criteria in a general system which any particular act of punishment must satisfy. One can, however, question the very idea of punishment as an institution that involves deliberately inflicting pain or deprivation. This raises the philosophical question of how one justifies a set of rules or an institution like a penal system. Corresponding to these two levels of justification are two broadly opposed approaches to punishment, the retributivist and the utilitarian. Each, in fact, has been taken to offer an answer to the problems at both levels, but the persuasive force of retributivism is mainly in its answers to problems of the first type, and of utilitarianism to questions of the second type. Characteristically, the retributivist stresses guilt and desert, looking back to the crime to justify punishment and denying that the consequences of punishment, beneficial or otherwise, have any relevance to justification. The utilitarian, on the other hand, insists that punishment can be justified only if it has beneficent consequences that outweigh the intrinsic evil of inflicting suffering on human beings.

**RETRIBUTIVIST THEORIES.** The most thoroughgoing retributivists, exemplified by Immanuel Kant, maintain that the punishment of crime is right in itself, that it is fitting that the guilty should suffer, and that justice, or the moral order, requires the institution of punishment. This, however, is not to justify punishment but, rather, to deny that it needs any justification. To say that something is

right or good in itself means that it does not need to be justified in terms of the value or rightness of anything else. Its intrinsic value is appreciated immediately or intuitively. But since at least some people do doubt that punishment is right, an appeal to intuition is necessarily unsatisfactory. Again, to say “it is fitting” or “justice demands” that the guilty should suffer is only to reaffirm that punishment is right, not to give grounds for thinking so.

Some retributivists, while admitting that punishment is, *prima facie*, evil, maintain that it is nevertheless better that the wicked should be punished than that they should prosper more than the virtuous and, perhaps, at their expense. In this view, the function of criminal law is to punish wickedness or immorality in order to maintain a kind of cosmic distributive justice. However, it is not self-evident that wickedness should be punished any more than it is self-evident that legal guilt should be. Archbishop Temple, himself a retributivist, declared that he had no “intuition that it is good that the wicked should suffer.” Nor is it clear that virtue must be rewarded or that universal justice requires the kind of human rectification that this sort of retributivism envisages. Of course, in a universe in which the wicked prospered, there might be no incentive to virtue, but this is essentially a utilitarian mode of argument. Again, evil motives and a bad character are necessary conditions of wickedness but not of legal guilt and criminal liability. The state’s function is to punish breaches of those rules which in the public interest ought to be upheld; it is a matter of indifference in law (but not in morals) that some men who observe the rules do so from the unworthy motive of fear and others break them from laudable motives of principle. Conversely, it is at least doubtful whether the criminal law should provide penalties for offenses against morality except where the public interest is at stake—for example, whether it should extend to cases of lying other than, say, false pretenses and perjury.

Though immorality is neither a necessary nor a sufficient condition for punishment, the relation between law and morals is nevertheless a close one, and what punishment is to the one, blame is to the other. Both regulate social intercourse, and in any given society the aims and ideals upheld by the law will usually correspond, more or less, with those upheld by the dominant morality. Moreover, in the family and the school punishment is often used to reinforce moral condemnation as part of the process of moral education. Some writers who regard punishment as moral retribution couple this idea with the argument that the point of punishment is to be found

in what Lord Justice Denning has called “the emphatic denunciation by the community of a crime.” In this view, punishment reinforces the community’s respect for its legal and moral standards, which criminal acts would tend to undermine if they were not solemnly denounced. There is, however, no intrinsic reason why denunciation should take precisely the form of inflicting suffering on criminals, unless, perhaps, one accepts Ewing’s view that punishment has the advantage of impressing both on the criminal and on everyone else that a breach of law and morals is so serious that society must do something to prevent it. That, however, is surely to justify punishment by its utility in maintaining respect for the law. Hastings Rashdall refers to “the enormous importance of the criminal law in promoting the moral education of the public mind,” but Rashdall was a utilitarian who justified punishment by reference to “the production of good effects on conscious beings.”

For G. W. F. Hegel punishment is necessary to annul the wrong done by the criminal. By this he means something more than restitution or compensation, neither of which is, strictly speaking, punishment. It is, rather, that the criminal has upset the balance of the moral order, which can be restored only by his being made to suffer. Or, in terms of the dialectic, crime is a negation of right and as such a nullity; punishment negates the negation, thus reaffirming the right. But in what sense can punishment be said to restore the balance or annul the wrong, unless it is taken for granted that criminals deserve to be punished? This is precisely the point in question.

**UTILITARIAN THEORIES.** The utilitarian position is exemplified in Jeremy Bentham’s remark that “all punishment is a mischief.... If it ought at all to be admitted, it ought only to be admitted in as far as it promises to exclude some greater evil.” By reforming the criminal, by deterring him or others from similar offenses in the future, or by directly preventing further offenses by imprisonment, deportation, or execution, the good that comes out of punishment may outweigh (so the utilitarian argues) the intrinsic evil of suffering deliberately inflicted. Without such effects, or if the suffering inflicted exceeded the suffering avoided, the institution would be unjustified.

The critics of utilitarianism claim that if people generally could be persuaded that an innocent man was guilty, utilitarianism would justify punishing him since as a warning to others he would be just as useful as a genuine offender. Again, offenders might be deterred by threatening to punish their wives and children, particu-

larly, if as is so often the case with political terrorists and resistance fighters, it were difficult to catch the offenders themselves. Or, again, if punishment could be justified as a way of reforming criminals, it would seem better to punish them before, rather than after, they committed their crimes. Retributivists claim that utilitarians are in danger of losing sight of two conditions that are necessary to the very idea of punishment—namely, that an offense should have been committed and that punishment shall be of the offender himself, who alone can be said to deserve it. “Punishment is punishment,” wrote F. H. Bradley, “only when it is deserved”; punishment for any other reason is “a crying injustice.”

The dilemma of utilitarianism, then, at least in its crude form, is that it justifies punishing innocent people provided that such punishment causes less suffering than might otherwise be caused by the would-be criminals it deters. Some utilitarians argue that in the end the deception would break down, that it could not be used systematically, or that the long-term consequences would be bad for society. But these answers are unsatisfactory because they depend on assumptions of purely contingent consequences. Our revulsion against punishing innocent people seems to go deeper than that. In any case, these answers will not meet the case for punishing hostages, which can certainly be done systematically and requires no deception or secrecy.

**PUNISHMENT AND PRINCIPLES OF JUSTICE.** To meet the above criticisms, a crude utilitarianism would have to be supplemented by other moral principles—namely, that differences in treatment must be justified by relevant differences in circumstance or condition, where “relevance” is defined in the light of general rules, and that every human being should be treated with at least a minimum of respect as a source of claims and not as a mere instrument for the promotion of the interests of others. It can be argued that punishment of the innocent or of hostages is an abuse not because it necessarily makes for more unhappiness than it prevents but because it treats innocent men in a way that is appropriate only for the guilty and makes an arbitrary difference in treatment between them and other innocent men. Moreover, a legal system is designed to guide conduct by laying down rules and attaching penalties to those who choose to break them. It is acceptable, in the words of J. D. Mabbott, only because “the criminal makes the essential choice; he ‘brings it on himself.’” Otherwise, punishment would not be consistent with the principle of respect for persons. The hostage, on the other hand, has no chance to settle his own fate; he is used as a mere lever for manipulating

other people's conduct, and his own interest is subordinate to that of the other members of society. Punishment of the innocent ignores, in short, fundamental procedural rules of justice and morality without which utilitarianism would make little sense, for unless everyone is worthy of equal consideration as a source of claims, whose interest is to count in assessing the utility of a course of action? Whom are we entitled to treat as simply a tool for advancing other men's interests—as Aristotle's "slave by nature"—and what would count as a reason for considering other men before him?

This has bearing, too, on the reasons for accepting as excuses such defenses as duress, unavoidable accident, or ignorance of fact—conditions under which an offender can claim that he could not help doing what he did. Bentham argued that to punish anyone under such conditions would be pointless and, therefore, mischievous, because the threat of penalties could not possibly deter anyone in the future who was similarly placed. Now, it is true that nothing would be lost if such people escaped punishment, provided they could be distinguished from cheats trying to take advantage of such excuses and provided enough offenders without such excuses could be detected to furnish examples for others. The principle of "strict liability," which exists in some legal systems for certain offenses, has been defended on the utilitarian ground that it is impossible to tell a genuine excuse from a pretense. It is questionable, however, whether a person who would otherwise be treated as innocent ought to be treated as guilty because someone else might otherwise escape a merited penalty. Punishing the man who commits an offense through ignorance or accident, because it is too difficult to tell whether he really did it on purpose or because we have to make an example of *someone*, is very like punishing the innocent as a warning to the guilty. The utilitarian case for these excuses is unsatisfactory inasmuch as it makes them subject to such qualifications.

A better ground for such excuses is that punishment is morally acceptable only if it is the consequence of an act freely chosen by the criminal, which it would not be under these conditions. A man acting in ignorance or by accident cannot be said to bring his punishment on himself. Punishment, seen as a way of influencing conduct, cannot be justified if there has been no real possibility of choice. Moreover, the punishment of involuntary offenses introduces into men's lives the possibility of disasters that they can neither foresee nor avert.

Utilitarianism, then, must be supplemented by principles of justice if it is not to clash with other moral prin-

ciples that are usually considered fundamental. It has, however, the merit, as an approach to the justification of punishment, that it provides a clear procedure for determining whether the institution is acceptable in general terms. This the retributivist approach cannot do because it denies the relevance of weighing advantages and disadvantages, which is what we ultimately must do in moral criticism of rules and institutions. Consequently, a retributivist justification of punishment as an institution usually turns out to be a denial of the necessity for justification, a veiled reference to the beneficial results of punishment (a utilitarianism in disguise), or an appeal to religious authority.

When it is a question of justifying a particular case of punishment, however, the retributivist is in a far stronger position. There would be no point in having a general rule if on every occasion that it had to be applied one had to consider whether the advantages in this particular case warranted acting in accordance with it. Moreover, the point of punishment as deterrent would be quite lost were there no general expectation, based on the general operation of the rule, that the guilty would be punished. Assuming, then, that a penal system can be justified in utilitarian terms, any offense is at least *prima facie* an occasion for a penalty. Equally, without an offense there is no question of a penalty. The retributivist contention that punishment is justified if, and only if, it is deserved is really applicable, therefore, to the justification of particular instances of punishment, the institution as such being taken for granted.

### SEVERITY OF PUNISHMENT

The clash between the utilitarian and retributivist approaches to punishment also arises in considering the criteria by which appropriate punishments are assessed. The retributivist insists that the punishment must fit the crime; the utilitarian relates the penalty to the general aims of the system, to the prevention of further crime, and, perhaps, to the reform of the criminal.

The most extreme form of retributivism is the law of retaliation: "an eye for an eye." This alone, Kant claimed, could provide a just measure of the penalty, since it was the crime itself and nothing else that settled it. However, to try to apply it literally might be monstrously cruel, or, as Kant recognized, it might be absurd. Thieves can be deprived of their property and murderers hanged, but what penalty is appropriate to the dope-peddler, the blackmailer, and the smuggler?

There is not much sense, either, in trying to construct a table of equivalents so that the amount of suffering

inflicted by the criminal could be meted out to him in some other form. How can such a table be drawn up? How many years must a blackmailer spend in jail to experience suffering equal to his victim's? Is it possible, in any case, to make comparisons of suffering between persons? Of course, we do assess the gravity of an offense and try to ensure that the punishment for a trivial offense is less severe than for a serious one. But this is possible only because we take for granted an existing scale of penalties and grade new offenses accordingly. Such grading does not imply an intrinsic relation between the crime and the penalty apart from that established by the scale. Some retributivists admit this but claim nevertheless that the penalties prescribed by the law ought to reflect the moral heinousness of the offense. The most serious offenses against morals deserve the most severe penalties. This, however, only shifts the question a step back, for what makes one moral offense more serious than another?

Utilitarians have tended to concentrate on deterrence, turning away from the actual criminal act except as one of a class of actions that might be prevented by punishing the particular instance severely enough (but only just enough) to make the action unattractive to the offender and to possible future offenders. Unfortunately, there are always people who cannot be deterred or reformed. Beyond a certain point the additional suffering one would have to inflict on all offenders to reduce their number might be so great as to exceed the amount of suffering thereby averted. The aim of the utilitarian, then, would presumably be to select the penalty at which the aggregate of suffering caused by crimes actually committed and punishments actually inflicted would be the smallest possible.

The utilitarian approach has often been criticized as justifying severe penalties for trivial offenses and vice versa. To eliminate parking offenses might need heavier penalties than to eliminate blackmail, which would be monstrous. But this criticism misses the point of the utilitarian case. There would, indeed, be no objection to threatening the severest penalty for any offense providing the threat never had to be carried out. Punishment is only an unfortunate consequence of the fact that the threats, which are the true operative elements in the system, are partially ineffective and would be wholly ineffective if they were not carried out when they failed to deter. In fixing penalties, the utilitarian's problem is not, therefore, to minimize the number of offenses, irrespective of the punishment inflicted, but to minimize the total amount of suffering from both sources. If we call parking offenses trivial, we mean that each one causes relatively little suf-

fering; therefore, we are prepared to put up with a large number of them rather than incur the cost of making offenders suffer heavy penalties. Blackmail, on the other hand, causes so much suffering that if heavier penalties would yield even a small reduction in the number of offenses, there might be a net gain even though offenders would suffer more than they did before. In this way a utilitarian might agree with the retributivist that severe penalties ought to be restricted to serious offenses, but he would argue that we call an offense serious precisely because it causes a great deal of suffering. For the retributivist only serious crimes *deserve* severe penalties; for the utilitarian only serious crimes are worth averting at the cost of severe penalties.

The utilitarian approach to this matter does not supply a procedure for sentencing particular criminals (any more than a justification for punishment as an institution would be a case for any particular application of it). Arguing from expected consequences, one might establish a kind of standard penalty for each class of offense. Officials drafting new rules might consider whether a proposed maximum penalty would keep offenses down to manageable proportions, or people concerned about road accidents might argue that heavier penalties for motoring offenses would make drivers more careful. Deciding the sentence in a particular case, however, is clearly a different matter. The maximum penalty is a limiting factor, but questions like the degree of responsibility, provocation, and the offender's previous record are all relevant. However, one might reasonably ask why, as a matter of principle, they should be relevant.

## PUNISHMENT AND RESPONSIBILITY

The problem of responsibility arises in relation to punishment as it does in relation to blame in moral theory. The principle, discussed already, that a man ought not to be punished for doing what he cannot help creates difficulties when extended to actions which a man could not help doing because of his own state of mind instead of external or contingent factors, like duress or ignorance of fact. An insane man, as defined, say, by the M'Naghten rules (that is, one who did not know what he was doing or did not know that what he was doing was wrong), cannot be said to choose his act because he cannot know it for what it is. But sometimes a man may know that what he is doing is wrong yet still be unable to stop himself from doing it. He may be subject, for instance, to an irresistible temptation or provocation. But how is that to be understood? A temptation is not irresistible merely because a particular man has yielded to it or even because

he might have been expected to yield to it. However, a temptation may be so strong that we might expect any ordinary person to yield to it (even though a few people may in fact resist it), or, as one might say, it might be “more than human nature can stand.” In that sense it may be “irresistible.”

Some people, of course, find it much more difficult than others to resist temptation. Some, like kleptomaniacs, are “impelled” to act in the sense that deliberation neither plays, nor could play, any part in what they do. Such people might be distinguished from plain wrongdoers by the fact that nothing—not blame, punishment, praise, or rational argument—seems to affect their disposition to break the rules. Or, again, their actions may lack any point, or if they can be said to have any point, it is only in relation to a set of aims and standards of achievement so distorted and eccentric that they are intelligible only to a psychiatrist. The kleptomaniac who steals nylon stockings for which he has no possible use (according to ordinary standards of utility) might properly be said to be unable to help stealing them. Far more difficult is the case of the psychopath, who seems to have no wish to resist temptation or, rather, who knows that some of the things he wants to do are wrong in the sense that other people disapprove of them but on whom this knowledge enforces no internal restraint beyond prompting a degree of caution. Criminals of this type would once have been described as “wicked” but are now often described as incapable of self-control. To say, however, that they are not responsible for their acts creates the odd situation that anyone is liable to punishment who usually resists temptation but sometimes fails, whereas the man who never resists is not liable at all.

The determinist has a short way with these difficulties. Since everyone’s actions are the response of his character to a given set of circumstances, how can anyone ever be held responsible for his actions? We do what we must, given what we are, and what we are is the end of a causal chain going back to before we were born. If one knew a person well enough, one might predict that under given conditions he would commit a crime. Is this compatible with saying that he can choose whether to do so, or is his belief in his freedom to choose simply an illusion? Can the result of a genuine choice be predicted?

To say that something is predictable is not, however, the same as saying it is unavoidable. We can forecast a man’s actions just because we know the kind of choices that he regularly makes. The more we know of his dispositions and his preferences, the more likely we are to be right. But that does not mean that he never acts voluntar-

ily or that he never makes a real choice but only thinks he does. If all choices are illusions, what would a real choice be like? A man’s behavior may be predictable because he can be relied upon to do what is reasonable, but to act with good reason is the very reverse of being subject to an inner compulsion. An essential difference between voluntary and involuntary action is that it makes sense to speak of the motives, aims, and reasons for the former but only of the causes of the latter. It is only when a person’s behavior seems pointless or when explanations in terms of aims do not seem sufficient that we look for the kind of cause which would justify saying that he could not help himself. Of course, a complete account of voluntary and rational behavior must refer to causes as necessary conditions for action, but such causes would not constitute a sufficient explanation. An account of the electronic activity in the brain would not provide a sufficient explanation of a move in a game of chess unless the move was so completely and absurdly irrelevant that it had to be accounted for simply as the result of a nervous twitch. In that case, however, it would not really be a move in the game at all, not an action, indeed, but something that happens to the player. The weakness of the determinist position, insofar as it purports to undermine the notion of responsibility, is that it treats such abnormalities as the explanatory model for the normal.

It is arguable, in any case, that the concept of responsibility *requires* that human behavior be causally accountable rather than the reverse. As David Hume pointed out in *An Enquiry concerning Human Understanding*,

[Where actions] proceed not from some *cause* in the character and disposition of the person who performed them, they can neither redound to his honour, if good; nor infamy, if evil. ... The person is not answerable for them; and as they proceeded from nothing in him that is durable and constant, and leave nothing of that nature behind them, it is impossible [that] he can, upon their account, become the object of punishment or vengeance.

In Hume’s view universal causality is consistent with the concept of choice and is a necessary condition for responsibility and, therefore, for blame and punishment.

Strictly speaking, all that is necessary for a theory of punishment is that human conduct should be capable of being modified by threats. For some people—for instance, compulsive lawbreakers like kleptomaniacs—that is not the case. Others, however, commit crimes believing they can escape punishment; still others, in a spirit of rebellion, indifference, or, more rarely, of mar-



tyrdom, prefer to do what they want and risk the consequences rather than conform. Why they prefer it—what conditions account for their being the men they are—is irrelevant. To say “they prefer it” is to say they might have chosen to do otherwise but did not, and that is all that is necessary for the concept “responsibility.” To ask whether they were free to prefer otherwise, being what they were, is to ask whether they could choose to choose, and it is not clear that this really means anything. The experience of punishment may provide a reason for choosing differently next time, but to have a reason for choosing is not to be without a choice and, therefore, without responsibility.

**EXTENUATION.** Though a criminal may be held responsible for his actions, there may nevertheless be circumstances which, so it is said, diminish responsibility or extenuate guilt. Temptation or provocation, though not irresistible may have been very great. The offender may have had a good character, and there may be no reason to expect any future lapse.

In some cases mitigation of sentence on such grounds can be readily justified in utilitarian terms. Little is to be gained by punishing the obviously exceptional lapse; a very small penalty might be enough to dissuade other respectable people who might otherwise be tempted to imitate it and for whom the shame of being treated as a criminal, whatever the penalty, is usually deterrent enough.

However, it is not easy to show, at least in utilitarian terms, that mitigation is reasonable in all the instances in which it is commonly thought appropriate. Nor does everyone agree on what are extenuating circumstances. It is not self-evident that whoever is sorely (but not irresistibly) tempted should be treated more leniently than people who have done the same thing but under less temptation. A strong temptation might be withstood if there were sufficient counterinducement. Leniency might weaken the resolve of others in the future. Some people treat crimes of passion leniently; others would say that the temptation is so commonly felt that if people were not discouraged from taking the law into their own hands by treating offenses of this kind severely, such offenses would rapidly multiply. Again, some people would accept a plea of drunkenness as an extenuation of an offense, whereas others would consider it an aggravation.

It is doubtful whether our ideas on this aspect of punishment depend on utilitarian considerations. Nor is there any reason to suppose that any system of utilitarian argument could show them to be consistent and rational.

It was suggested earlier that though the criteria of morality and law, of blame and punishment, are not identical, they influence one another. If we blame people less for yielding to strong temptation, we also feel they deserve a less severe punishment. But this only shifts the question a step back. Why should temptation mitigate blame?

A possible answer might be that at least some temptations can be pleaded as partial justifications. Thus, a man who pleads that he killed someone to shorten his sufferings or a woman who kills her deformed baby is appealing to another moral principle to excuse the act. Similarly, a man who kills his wife’s lover might claim that his victim was violating his rights. These are not complete justifications, as a plea of self-defense would be, but they are excuses that count, as it were, against the initial presumption of guilt and so incline us to look at the offense more sympathetically and more leniently, whatever the advantages of severity in terms of deterrence, prevention, or reform. There is nothing irrational in striking a balance of desert.

But differences of opinion about a criminal’s deserts often turn not on the way such a balance is struck but on the extent to which his judges (or their critics) are able to comprehend his action. Anyone who could imagine himself tempted in similar circumstances would probably be more sympathetic than someone who could not and who would therefore see no reason for being indulgent. On the other hand, anyone who suspected that he himself might yield to such a temptation and who flinched from the possibility might react to it with very great severity indeed.

## PUNISHMENT AND REFORM

There is no reason to suppose, then, that the sentencing practice of the courts will display rational and consistent principles; furthermore, any attempt to set up criteria of rational judgment on strictly utilitarian principles is likely to cut across deeply rooted moral convictions. Accordingly, some criminologists and psychiatrists, such as Eliot Slater and Bernard Glueck, and some penal reformers, such as Barbara Wootton, have swung away from the general conceptions of punishment and desert. Instead of asking what penalty is warranted by the crime, whether the agent was fully responsible for his action, whether circumstances exonerate him wholly or in part, they prefer to ask what kind of treatment is most likely to rehabilitate him, subject, of course, to the example it might set for others.

This comes very close to repudiating altogether the concept of punishment as a deliberate infliction of suffer-

ing, which the criminal deserves, consequent to a voluntary breach of the law. First, the treatment most likely to rehabilitate him need not be unpleasant (though if it is to instill a measure of discipline, it very well may be). And, second, avoiding the question of moral responsibility, the reformer also avoids the question of what the criminal deserves, because the reformer's prime concern is with the treatment he needs. Criminals would no more deserve punishment than the sick deserve medicine. Indeed, for such writers as Samuel Butler and the American lawyer Clarence Darrow, criminality is a kind of sickness to be treated rather than a wrong to be punished.

Attractive as this approach may seem on humanitarian grounds, it has at least one serious consequence. The concepts of responsibility and desert cannot be discarded without some loss. For it is not a necessary condition of medical treatment that a patient must have shown symptoms of a disease; those exposed to smallpox are vaccinated before they develop a fever. Without the principle that punishment must be deserved, there would be no obstacle to subjecting people likely to become criminals to corresponding forms of penal prophylaxis. Moreover, if we substitute for punishment the idea of rehabilitative treatment, there is nothing against sentencing a person of bad character to a severe course of treatment for the most trivial offense if his character would be better for it in the end. This would clearly be incompatible with the usually accepted principle that trivial offenses should not carry severe penalties.

Reformism of this kind is open to attack from another quarter. The point has been made by Hegel and Bernard Bosanquet, among others, that retributive punishment is a kind of tribute to the moral personality of the criminal. It is precisely as a morally responsible agent, recognized as capable of making reasoned choices and accepting the consequences, that the criminal is punishable. Bosanquet goes so far as to say that punishment is "his right, of which he must not be defrauded." It is to be distinguished, argued Bradley, from the discipline or correction appropriately administered to animals and children. Punishment "is inflicted because of wrongdoing, as desert, the latter is applied as means of improvement." Since rational adults are neither animals nor children, no one has the right to treat them as if they were. It might be similarly argued that lunatics are under tutelage because they are incapable of looking after their own interests and cannot be expected to respect those of other people. The sane criminal, on the contrary, can be made to pay for his antisocial choices in order to demonstrate to him and, through him, to others that crime does not pay, but it

diminishes his stature as a rational adult to deny that he is responsible for ordering his own life and to impose upon him ends of another person's choosing.

Nevertheless, retributivists have often been much concerned with moral reformation. They have insisted, however, that this was something the criminal must do for himself. Because it was associated with shame and rejection, punishment could bring the criminal up short and force him to reconsider his life in the light of society's condemnation of his actions. But the remorse that was a necessary condition for self-reformation was entirely dependent on the criminal's recognition that his punishment was deserved. Without that there could be no inward reformation, no reassertion of moral standards, but only a sense of resentment and injustice. Accordingly, punishment can yield the benefits of reform only if it is thought of, above everything else, as retributive—as the appropriate desert of a responsible guilty agent. It is this which distinguishes the retributive approach to moral reformation from the kind of utilitarianism which turns its back on desert and responsibility and is concerned only with the needs of rehabilitation.

It is, of course, an open question whether punishment ever does produce the kind of self-reformation the Hegelians had in mind or whether it does so more often than it produces a moral decay. Indeed, our knowledge of the facts of criminal behavior is probably far too scanty and uncertain for us to know how relevant much of the philosophical discussion of punishment really is. We cannot say for sure that a penal system is justified because it tends to reform criminals. Nor do we know, for that matter, whether the deterrent view of punishment is applicable to all kinds of crime. Many people commit offenses without seeming to take any account of consequences before they act, and they repeat the same offenses again and again in spite of punishment. Perhaps those who do not, would not repeat them even without punishment. Perhaps there would be no more cases of certain classes of crime than there are already; perhaps the only people to commit them are those who also do not take account of consequences before they act. It seems likely that some potential offenders are deterred from evading taxes or from smuggling by the threat of punishment, but is there any certain evidence that the threat of punishment deters anyone who would otherwise commit rape or arson?

Utilitarians tend to assume that punishment as an institution can be justified by its beneficial consequences, but the argument depends on certain a priori assumptions about criminal (or would-be criminal) behavior that may be greatly overintellectualized. However, even though

research should prove the usual utilitarian justifications for punishment groundless, that does not mean that some other, nonutilitarian justification is better. The proper procedure may well be to ask, with the utilitarian, whether the consequences are by and large beneficial; it is equally possible that punishment as an institution might fail that test. A theory of punishment that led to the conclusion that all punishment was wrong need be no more necessarily mistaken than a theory that led to a similar conclusion as regards, say, slavery, which, after all, was accepted as uncritically in Aristotle's day as punishment is today.

**See also** Aristotle; Bosanquet, Bernard; Bradley, Francis Herbert; Butler, Samuel; Good, The; Hegel, Georg Wilhelm Friedrich; Hume, David; Kant, Immanuel; Moral Rules and Principles; Rashdall, Hastings; Responsibility, Moral and Legal; Utilitarianism.

### Bibliography

For a broad treatment of the subject, see A. C. Ewing, *The Morality of Punishment* (London: K. Paul, Trench, Trubner, 1929). For classic expositions of retributive positions, see Immanuel Kant, "Metaphysische Anfangsgründe der Rechtslehre," in *Die Metaphysik der Sitten* (1797), in *Werke*, Vol. VII (Berlin, 1922), translated by W. Hastie as *Kant's Philosophy of Law* (Edinburgh, 1887); G. W. F. Hegel, *Philosophie des Rechts* (1821), in *Werke*, Vol. VIII (Berlin, 1833), translated by T. M. Knox as *Hegel's Philosophy of Right* (Oxford: Clarendon, 1942). See also Ch. 8 of Bernard Bosanquet, *Philosophical Theory of the State* (London, 1899; 4th ed., Macmillan, 1923), and Essay 1 of F. H. Bradley, *Ethical Studies* (London, 1876; 2nd ed., rev., Oxford: Oxford University Press, 1927). Extensive bibliographic notes can be found in Giorgio Del Vecchio, *La giustizia*, which was originally published in *Annuario dell'Università di Roma* (1922–1923), 4th ed. (Rome, 1951), translated by Lady Guthrie as *Justice*, edited by A. H. Campbell (Edinburgh, 1952).

More recent analyses, mainly in retributive terms, can be found in J. D. Mabbott, "Punishment," in *Mind* 48 (1939): 152–167; C. W. K. Mundle, "Punishment and Desert," in *Philosophical Quarterly* 4 (16) (1954): 216–228; D. C. Hodges, "Punishment," in *Philosophy and Phenomenological Research* 18 (2) (1957–1958): 209 (a case for viewing punishment as reparation). K. Baier, "Is Punishment Retributive?," in *Analysis* 16 (2) (1955): 25–32, argues that the concept of punishment entails retribution but does not deal with problems of justification. See also A. R. Manser, "It Serves You Right," in *Philosophy* 37 (142) (1962): 293–306. H. J. McCloskey, "The Complexity of the Concepts of Punishment," *ibid.*, 307–326, draws attention to variety of contexts in which "punishment" is used and argues against analysis in terms of a standard case.

For classic statements of utilitarian positions, see C. B. Beccaria, *Dei delitti e delle pene* (1764); *An Essay on Crimes and Punishment, Translated from the Italian*, 4th ed. (London, 1785); Jeremy Bentham, *Introduction to the*

*Principles of Morals and Legislation* (1789), Wilfrid Harrison, ed. (Oxford, 1948), and *Traité de législation, civile et pénale*, E. Dumont, ed. (Paris, 1802), translated by R. Hildreth as *Theory of Legislation*, C. K. Ogden, ed. (London, 1931). See T. H. Green, *Lectures on the Principles of Political Obligation*, in *Works*, edited by R. L. Nettleship, Vol. II (London, 1885), or in a separate edition, with an introduction by A. D. Lindsay (London: Longman, 1941); Green repudiates retributivism but stresses moral reformation as a function of punishment and, therefore, the importance of the concept of desert (cf. Bosanquet, *op. cit.*). See also Vol. I of Hastings Rashdall, *Theory of Good and Evil*, 2 vols. (Oxford: Clarendon Press, 1907; 2nd ed., 1924).

The following examples of recent analyses of the problems of punishment are mainly utilitarian in emphasis but try to allow some place for retributivism in a general theory. In W. D. Ross, *The Right and the Good* (Oxford: Clarendon Press, 1930), especially Appendix II, Ross argues that the state is entitled to punish offenders in the public interest because by invading the rights of others, they forfeit their own claim to protection. On this subject see also A. G. N. Flew, "The Justification of Punishment," in *Philosophy* 29 (3) (1954): 291–307; H. L. A. Hart, "Prolegomenon to the Principles of Punishment," in *PAS* 60 (1959–1960): 1, and "Murder and the Principles of Punishment: England and the United States," in *Northwestern University Law Review* 52 (4) (1957): 433. Also, see A. Quinton, "On Punishment," and Stanley I. Benn, "An Approach to the Problems of Punishment," in *Freedom and Responsibility*, edited by Herbert Morris (Stanford, CA: Stanford University Press, 1961); J. Rawls, "Two Concepts of Rules," in *Society, Law, and Morality*, edited by F. A. Olafson (Englewood Cliffs, NJ: Prentice-Hall, 1961); Stanley I. Benn and R. S. Peters, *Social Principles and the Democratic State* (London: Allen and Unwin, 1959), especially Chs. 8–9. Patrick J. Fitzgerald, *Criminal Law and Punishment* (Oxford: Clarendon Press, 1962), deals with the philosophical problems in the broad context of English criminal law. For a discussion of the relations of law, morality, and punishment, see Patrick Devlin, *The Enforcement of Morals* (London: Oxford University Press, 1965), and H. L. A. Hart, *Law, Liberty and Morality* (Stanford, CA: Stanford University Press, 1963).

For discussions of free will, determinism, and responsibility in relation to punishment, see H. L. A. Hart, Paul Edwards, and John Hospers, symposium on "Determinism and Freedom in Law and Ethics," in *Determinism and Freedom*, edited by Sidney Hook (New York: New York University Press, 1958); J. D. Mabbott, "Freewill and Punishment," in *Contemporary British Philosophy*, edited by H. D. Lewis, 3r series (London, 1956). In Barbara Wootton, *Social Science and Social Pathology* (London: Allen and Unwin, 1959), there is a survey of recent trends in psychiatric and criminological studies critical of the concept "responsibility"; see especially Ch. 8. For a criticism of these trends, see H. L. A. Hart, *Punishment and the Elimination of Responsibility* (Oxford: Clarendon Press, 1968). *Freedom and Responsibility* (see above) is a useful collection of readings on various aspects of the problem of responsibility, including its relation to punishment, extenuation, and so on; it includes an extensive bibliography.

**Stanley I. Benn (1967)**

## PUNISHMENT [ADDENDUM]

Since 1967, preventive theories of punishment (whether strictly utilitarian or more loosely consequentialist) have entered a long decline, beginning with the virtual disappearance of reform theory in the 1970s. Crowding them out are various alternatives generally categorized as “retributive.”

All preventive theories treat punishment as (primarily) a means of controlling objectionable behavior. Insofar as they propose to justify punishment on the assumption that penalties can be tuned to achieve a certain degree of social control, they are empirically vulnerable (as well as morally vulnerable for ignoring justice). What had become clear by the 1970s was that social science could not then, or in the foreseeable future, give preventive theories much empirical content. Social science could not, that is, say what effect, if any, statutory penalties, rehabilitation, exemplary punishments, or even the incapacitation of criminals would have on the crime rate. If even relatively crude tuning of penalties to consequences is in practice impossible, preventive theories cannot justify choosing one system of punishment over another, much less one punishment over another.

In contrast to preventive theories, retributive theories do not seek to justify punishment by pointing to an empirical relation between punishment and “consequences” (such as a certain crime rate). For retributive theories, the relation between punishment and its justification is conceptual (“internal,” as Georg Wilhelm Friedrich Hegel would say). No retributivist need deny that punishment has some general tendency to control crime or that that tendency is a reason to have some punishment system rather than none. All retributivists need deny is that punishment’s (actual or probable) tendency to control crime matters much for understanding why we should (or should not) have this institution, practice, or act of punishment rather than another.

Retributive theories may be divided into two importantly different kinds: moralistic and legalistic. Moralistic retributivism has three (main) divisions: desert, paternalist, and condemnatory theories. *Desert* theory takes it as (more or less) brute fact that wrongdoing deserves an unpleasant response, that is to say, punishment. Punishment is justified because it is deserved. Nothing more need be said. *Paternalist* theory holds that all justified punishment, or at least all justified punishment of rational agents, must aim at a certain good for those punished. This good may be subjective (R. A. Duff’s

“penance”) or objective (Robert Nozick’s “connection with correct values”). *Condemnatory* theory, in contrast, understands punishment as (primarily) an “expressive act” not meant to benefit anyone.

All three varieties of moralistic theory are retributive (in the sense used here) because all seek to achieve a good that is conceptually related to punishment. For desert theories, that good is simply giving wrongdoers what they deserve. Degree of desert determines severity of punishment. For paternalist theories, the justification of punishment lies in the way punishment treats the wrongdoer—for example, as a being capable of learning justice from the punishment appropriate to the crime. The seriousness of the wrong determines what penalty is appropriate to teach the lesson that the crime shows the wrongdoer needs to learn. For condemnatory theories, the justification of punishment lies in what the punishment “expresses.” The denunciation should be as emphatic as the crime was bad; the more severe the punishment, the more emphatic the denunciation is.

While desert theories seem to be the direct descendants of traditional retributivism, paternalist theories superficially resemble traditional *reform* theories, and condemnatory theories similarly resemble traditional *deterrence* theories (denunciation resembling a deterrent threat). Both nonetheless differ fundamentally from the corresponding preventive theory. According to the paternalist theory (in its pure form, at least), punishment would be justified even if wrongdoers never repent or learn as a result of punishment. What is important—important because it respects the moral personality of the wrongdoer—is that the right punishment be imposed with the right intention. In much the same way, according to the condemnatory theory (in its pure form), punishment is justified even if emphatic denunciation has no effect on the crime rate or on the individual’s later conduct. Reaffirming the wrongness of an act is good in itself, good enough (all else being equal) to justify the punishment.

All moralistic theories share the assumption that punishment belongs to ordinary morality (rather than to the law in particular). Moralistic theories use ordinary moral practices (such as disciplining children) to understand punishment (with legal punishment only a special case). Moralistic theories differ primarily in the part of ordinary morality to which they assign punishment. Desert theory treats punishment as (negative) rewarding; paternalist theory treats punishment as correction or teaching; and condemnatory theory treats punishment as moral statement. Other moralistic theories are possible,

for example, one treating punishment as a form of self-defense or satisfaction of a promise.

Legalistic theories, in contrast, assume that (justified) punishment is a practice (largely) confined to (relatively just) legal systems. The only important form of legalistic retributivism today is “the fairness theory” (also known as “benefits-and-burdens,” “reciprocity,” or “unfair advantage” theory). It holds that legal punishment (and close analogues) is justified insofar as it supports the (relatively) just distribution of benefits and burdens that a (relatively just) legal system (or similar practice) creates. A relatively just legal system is to be thought of as a cooperative enterprise from which each benefits if others generally do their part and in which doing one’s part will sometimes be burdensome. According to the fairness theory (in its pure form at least), the institution of legal punishment is justified if punishment keeps lawbreakers from gaining an unfair advantage over the law-abiding. Punishment, if just, necessarily takes back the unfair advantage the crime as such takes (or, at least, some fair equivalent of that advantage). Though the fairness theory has an obvious affinity to certain theories of distributive justice (especially, Rawlsian social contract), it presupposes no particular theory. All it presupposes is that there can be an equivalence between crime and (just) punishment assuring that (in general at least) legal punishment of certain people in certain ways will (as a conceptual matter) increase (or at least help to maintain) overall distributive justice (however defined). Explaining that presupposition has proved difficult.

**See also** Hegel, Georg Wilhelm Friedrich; Nozick, Robert; Presupposition; Rawls, John.

### Bibliography

For some sense of the history of punishment since 1967, compare Marvin Zalman, “The Rise and Fall of the Indeterminate Sentence,” *Wayne Law Review* 24 (1977): 45–94; Nigel Walker, *Punishment, Danger, and Stigma* (Oxford: Blackwell, 1980); and Annika Snare, ed. *Beware of Punishment*, *Scandinavian Studies in Criminology*, 14 (1995).

For some broad general treatments of punishment theory, see Ted Honderich *Punishment: The Supposed Justifications* (London: Hutchinson, 1969); Hyman Gross, *A Theory of Criminal Justice* (New York: Oxford University Press, 1979); C. L. Ten, *Crime, Guilt, and Punishment: A Philosophical Introduction* (Oxford: Clarendon, 1987); Igor Primoratz, *Justifying Legal Punishment* (Atlantic Highlands, NJ: Humanities Press International, 1989); and Michael Moore, *Placing Blame: A General Theory of the Criminal Law* (Oxford: Clarendon, 1997). Comparison of these works (spaced about a decade apart) suggests the increasing

importance of retributivism, even when (as with Ten) the author himself argues for a preventive theory.

For some important criticism of retributivism, see Richard W. Burgh, “Do the Guilty Deserve Punishment?” *Journal of Philosophy* 79 (1982): 193–210; John Cottingham, “Varieties of Retribution,” *Philosophical Quarterly* 29 (1979): 236–246; John Braithwaite and Philip Pettit, *Not Just Deserts: A Republican Theory of Criminal Justice* (Oxford: Clarendon, 1990), ch. 8–9; David Dolinko, “Some Thoughts about Retributivism,” *Ethics* 101 (1991): 537–559; David Dolinko, “Retributivism, Consequentialism, and the Intrinsic Goodness of Punishment,” *Law and Philosophy* 16 (1997): 507–528; and Nigel Walker, “Even More Varieties of Retribution,” *Philosophy* 74 (1999): 595–605.

For the best recent statement of the desert version of retributivism, see D. E. Scheid, “Constructing a Theory of Punishment, Desert, and Distribution of Penalties,” *Canadian Journal of Law and Jurisprudence* 10 (1997): 441–506. Other important discussions include Lawrence H. Davis, “They Deserve to Suffer,” *Analysis* 32 (1972): 136–140; John Kleinig, *Punishment and Desert* (The Hague: Nijhoff, 1973); W. A. Parent, “The Whole Life View of Criminal Desert,” *Ethics* 86 (1975): 350–354; and Douglas N. Husak, “Why Punish the Deserving?” *Nous* 26 (1992): 447–464.

For the most influential statements of the paternalistic theory, see Herbert Morris, “A Paternalistic Theory of Punishment,” *American Philosophical Quarterly* 18 (October 1981): 263–271; Robert Nozick, *Philosophical Explanations* (Cambridge, MA: Harvard University Press, 1981), pp. 363–397; Jean Hampton, “The Moral Education Theory of Punishment,” *Philosophy and Public Affairs* 13 (1984): 208–238; and R. A. Duff, *Trials and Punishments* (Cambridge, U.K.: Cambridge University Press, 1986).

Among important discussions of the expressive theory are: A. J. Skillen, “How to Say Things with Walls,” *Philosophy* 55 (1980): 509–523; Igor Primoratz, “Punishment as Language,” *Philosophy* 64 (1989): 187–205; Michael Davis, “Punishment as Language: Misleading Analogy for Desert Theorists,” *Law and Philosophy* 10 (1991): 310–322; Andrew von Hirsch, *Censure and Sanctions* (Oxford: Clarendon, 1993). Though Joel Feinberg’s influential “The Expressive Function of Punishment,” *Monist* 49 (1965): 397–423, is usually treated as a seminal statement of the expressive theory—as reprinted in Feinberg’s *Doing and Deserving: Essays in the Theory of Responsibility* (Princeton, NJ: Princeton University Press, 1970)—it is not. It actually proposes to revise the standard (Hart-Flew) definition of punishment, something necessary because of Feinberg’s legal positivism. Since retributivism aims at justice, not mere social control, retributive expressivists do not need to fiddle with the definition of punishment to distinguish between (justified and so, expressive) punishment and mere penalties. Retributivists have therefore generally ignored Feinberg’s proposal (while citing him as pointing to the expressive function of punishment).

For some examples of moralistic retributivism relying neither on desert, paternalism, nor condemnation, see: Phillip Montague, *Punishment as Societal Defense* (Lanham, MD: Rowman & Littlefield, 1995); and Jerry Cederblom, “The Retributive Liability Theory of Punishment,” *Public Affairs Quarterly* (1995): 305–315.

For important discussions of the fairness theory (and related issues), see: Herbert Morris, "Persons and Punishment," *The Monist* 52 (1968): 475–501; John Finnis, "The Restoration of Retribution," *Analysis* 32 (1971–1972): 131–135; J. Murphy, "Marxism and Retribution," *Philosophy and Public Affairs* 2 (1973): 217–243; Wojciech Sadurski, *Giving Desert its Due: Social Justice and Legal Theory* (Dordrecht: Kluwer, 1985); George Sher, *Desert* (Princeton, NJ: Princeton University Press, 1987); Richard Dagger, "Playing Fair with Punishment," *Ethics* 103 (1993): 473–488; and Anthony Ellis, "Punishment and the Principle of Fair Play," *Utilitas* (1997): 81–97. For an example of a legalist retributivism that at least seems not to be an instance of the fairness theory, see Margaret R. Holmgren, "Punishment as Restitution: The Rights of the Community," *Criminal Justice Ethics* (1983): 36–49.

For recent examples of the application of retributivism to various practical problems in the theory of punishment, see Kathleen Dean Moore, *Pardons: Justice, Mercy, and the Public Interest* (New York: Oxford University Press, 1989); Jeffrie G. Murphy and Jean Hampton, *Forgiveness and Mercy* (Cambridge, U.K.: Cambridge University Press, 1990); Michael Davis, *To Make the Punishment Fit the Crime: Essays in the Theory of Criminal Justice* (Boulder, CO: Westview, 1992); R. A. Duff, *Criminal Attempts* (Cambridge, U.K.: Cambridge University Press, 1996); Andrew von Hirsch and Andrew Ashworth, eds. *Principled Sentencing: Readings on Theory and Policy* (Oxford: Hart, 1998); and Stephen Kershnar, "An Argument for the Use of Torture as Punishment," *Hamline Journal of Public Law and Policy* (1998): 497–528.

Recent positive work on deterrence seems to be more than a decade old. See, for example (beside Ten, 1987, above), D. Schmitz, "Deterrence and Criminal Attempts," *Canadian Journal of Philosophy* 17 (1987): 615–623; Daniel M. Farrell, "On Threats and Deterrence," *Social Theory and Practice* 15 (1989): 367–394; and B. Sharon Bird, "Kant's Theory of Punishment: Deterrence in its Threat, Retribution in its Execution," *Law and Philosophy* 8 (1989): 151–200. Reform theory may be reviving as "restorative justice" within the victim's rights movement (but with a strong retributive overlay). See, for example, Charles K. B. Barton, *Getting Even: Revenge as a Form of Justice* (Chicago: Open Court, 1999).

*Michael Davis (2005)*

## PUTNAM, HILARY

(1926–)

Hilary Putnam, after receiving a BA from Pennsylvania (1948) and a year spent at Harvard (1948–1949), studied at the University of California, Los Angeles, taking his doctorate in 1951 with a dissertation titled "The Concept of Probability: An Application to Finite Sequences." He taught at Northwestern (1952–1953), Princeton (1953–1961) and MIT (1961–1965), becoming Walter Beverly Pearson professor at Harvard in 1965. From 1995 to 2000,

he served as Cogan University Professor there, becoming emeritus in 2000. He has been influential in most areas of philosophy, particularly in the philosophy of language, of logic, of mathematics, and of science.

Putnam is sometimes thought of as often changing his mind. (See, for example, the *Dictionary of Philosophers' Names*.) Sometimes he has. But in central respects he has held a single, though developing, position since the mid-1950s, a position that in some aspects resembles the later Ludwig Wittgenstein's. This entry sets out some constant central themes.

Putnam was among those American philosophers to benefit directly from the intellectual exodus from Europe caused by Nazism. He was a student of Rudolf Carnap and of Hans Reichenbach. Though his approach to issues is quite different from theirs, Reichenbach in particular had a lasting and often acknowledged influence on Putnam's thought. Putnam's innovations stand out when it is noted. In *Realism with a Human Face* (1990, p. 289), Putnam remarks,

In *Theory of Relativity and A Priori Knowledge* (1922) Reichenbach listed a number of statements ... each of which Kant would have regarded as synthetic *a priori*, and each of which can be held immune from revision ..., but which *collectively imply statements that are empirically testable*, and that Kant would, therefore, have to regard as *a posteriori*.

Certain principles had, in Immanuel Kant's time, as good a claim as any to fix how particular spatial, temporal, and other concepts are to be applied, and thereby which concepts those were; to be intrinsic to the concepts involved, thus "conceptual truths," thus *a priori*. Relativity theory allows us to see how they are at least jointly testable, so that some may turn out false. Such, it seems, is a fate to which *a priori* truths are liable.

Putnam reports Reichenbach as making a related point to his classes. Considering questions such as "How can we show that that blackboard is wider than this ash-tray?" he argued that any system of measurement, or of observation, treats some propositions that seem empirical (such as "mere translation does not make things grow or shrink") as axiomatic. One cannot sensibly apply the system while doubting these propositions; they are not subject to confirmation or refutation within the system. But it could prove reasonable to replace the system with another in which these propositions are testable, so possibly false. In that sense they are empirical.

There are two contrasting reactions to these points. One is: What this shows is that every concept commits itself to a particular empirical theory. If the theory proves false, then the concept is incoherent, so without application and to be discarded. This was Paul Feyerabend's reaction, and it is also Paul Churchland's.

The other reaction is: If we are confronted with situations that force giving up what seemed conceptual truths, it may appear that the concepts whose applications seemed to be governed by those principles are, in fact, otherwise governed. Perhaps the application of the concept "straight line" to items in the world is not governed by the Euclidean parallel postulate, but rather in such and such other way. That reaction grants face value to Reichenbach's point that the same proposition that is axiomatic in one system may be testable and false in another. This was Putnam's reaction. He developed the position and drew its implications in a powerful series of papers in the 1950s and 1960s (see Putnam, 1962a–d). Part of the idea is that what principles govern the application of a concept depends in part on how the world in fact is. Putnam defined that role for the world in "The Meaning of 'Meaning'" (1975). This last article, though not published until 1975, was completed by 1968.

By the early 1970s, Putnam had begun to emphasize some new themes. For one thing, he became increasingly impressed with what he calls the "interest relativity" of such notions as explanation and cause. The general point is: What a concept counts as applying to—the correct way of applying it—varies with the circumstances in which it is to be applied. A concept may count, on one occasion, as fitting what it does not count as fitting on another. That is continuous with Putnam's earlier reaction to Reichenbach. The point then was: What it is reasonable to judge as to how a concept operates depends on the conditions in which such judgments are made. The point now is: What those conditions are depends not just on how the world is, but may vary from occasion to occasion, given the world as it is. Not coincidentally, this point went along with other developments in Putnam's thought.

The first of these developments is what he calls "internal realism," first presented in 1976, and amplified in his writings of 1981, 1983, 1987a, 1987b, and elsewhere. The position includes four points. First, there are mundane, true things to say about what our words and thoughts are about: "the word 'gold' means (refers to) gold; this is gold"; "This is a chair; this is what 'chair' refers to," and so forth. Second, there are philosophical dicta that sound much like such mundanities, or their denials, or generalizations of these, but that say, or try to

say, something quite different. They are bad answers to the following pseudo-problem. On the one hand, there are thoughts and words—items that purport to represent the world as being thus and so; on the other hand, the items the world in fact contains, which are what and how they are independent of what we think, or do not think, about them. How are our words and thoughts related to these items? How, if at all, does their truth depend on how those items are? And how could they be so related? Internal realism holds that the problem rests on a mistake; hence so do any 'solutions', which take it at face value.

Third, the mundane remarks (point one) are correct because they are a feature of how these words are (or are to be) used. But that formulation depersonalizes things misleadingly. The standard for the correctness of a statement cannot be fixed independently of what users of the relevant words and concepts—that is, human beings—are prepared to recognize as correct: What Putnam identifies as our (human) perceptions of rationality and reasonableness. What it is for a statement to be correct depends on the sorts of beings we are, and is not reducible to some set of principles that would have to hold anyway. Fourth, it is part of what we are prepared to recognize as rational that any concept might be applied correctly in different ways in different circumstances. What sometimes counts as the cause of the explosion may not at other times. It is because human rationality is occasion-sensitive that the problem mentioned in point two is a pseudo-problem. We cannot sensibly take a "God's-eye view" of how we relate to the world, trying to say how our concepts would apply without us.

The occasion-sensitivity of rationality does not mean that truth is relative, or that there are no objective facts—given a framework, or setting, in which concepts are to be applied. Nor does giving up on a God's-eye view mean a deflationist account of truth. Putnam insists that we cannot comprehend what truth is without understanding the role of truth in our lives, notably in our activities of asserting, and of treating assertions in the ways we do; and that deflationism does not help us understand the role of truth in human life.

In arguing against the possibility of a God's-eye view, Putnam has produced what are probably his most discussed arguments. In one he identifies the God's-eye view (what John McDowell has called "the view of the cosmic alien") as one from which we may consider our own language as an uninterpreted calculus with a range of possible interpretations, and then ask which interpretation is the right one. In what he first saw as a generalization of the Skolem paradox, he argues that, in that case, *nothing*

could make one interpretation the right one, so we could not ever be talking about anything (or about one thing rather than others). But we cannot pose serious problems without talking about definite things. This is a reductio of the idea of a God's-eye view (see "Models and Reality," *Philosophical Papers*, Vol. 3).

In another argument Putnam considers the (apparent) question whether we might be brains in a vat: that we are, and always have been, nothing but brains, kept alive by a bath of nutrients, fed computer-generated stimuli through electrodes. He argues that if the God's-eye view is possible, then that we are, and that we are not such brains should both be possibilities. But for the words of the question to mean what a God's-eye view requires them to mean we must be using them in ways that entail that we are not brains in vats. For, as argued in "The Meaning of 'Meaning,'" what our words mean depends, *inter alia*, on how we are in fact connected to the world, and not just on what we may anyway be aware of. For our "brain" to mean *brain*, and our "vat" to mean *vat*, we must be connected to the world as brains in ways vats could not be. So we cannot formulate what, from a God's-eye view, ought to be a possibility, in a way that makes it possible. That is another reductio of the idea of a God's-eye view (see Putnam 1981).

Equally important to internal realism are Putnam's arguments against a causal theory of reference: Arguments, based on the interest-relativity of causation, that our being causally linked to the world as we, in fact, are is not enough in itself to make some one interpretation of our language correct—once it is granted that the language we speak may coherently be viewed by us as less than fully interpreted, so open to interpretation. These arguments appear in many replies to critics, and notably in "Realism with a Human Face" (1990).

At about the time Putnam began to develop internal realism, he also began to change his way of thinking about human psychology, rejecting a picture of it, and with that, a view he once espoused—functionalism. Viewed one way, a human being is an organism constructed in a particular way, a particular battery of mechanisms arranged to interact with each other and the environment in given ways. If, while taking that view, we ask what it is for someone to believe that Mars is a planet, or to have any propositional or other attitude—to be in a mood, experience an emotion, and so on—it is tempting to look for an answer by trying to identify some state(s) of some mechanism(s) such that for someone to believe that is for him to be structured like that. In that frame of mind, for example, one might speak seriously of someone

having a "token of a mentalese sentence" in his "belief box." This is the picture Putnam rejects.

Against it Putnam notes that to ascribe belief to someone is to relate that person to the world as we view it, and to ourselves, as on the same side as ours, or a different one, with respect to such and such question as to how the world is, and so on. Given internal realism, this means that there will be different truths to tell on different occasions as to what a given person, as he is at a given time, then believes. So for someone to be as said to be when we say him to believe thus and so, cannot be for him to have some particular mechanism, otherwise identifiable, in some particular state. And so on for other mental states (see Putnam 1989).

Putnam has been refining the ideas discussed above, notably the idea of a distinction between ordinary and philosophic statements, and applying them in new areas, such as philosophy of mathematics. The above indicates a few main themes, omitting Putnam's striking arguments for them.

**See also** Carnap, Rudolf; Functionalism; Kant, Immanuel; Rationality; Realism; Reference; Reichenbach, Hans; Wittgenstein, Ludwig Josef Johann.

## Bibliography

### WORKS BY PUTNAM

- "The Analytic and the Synthetic" (1962a). Reprinted in *Mind, Language and Reality, Philosophical Papers*, Vol. 2. Cambridge, U.K.: Cambridge University Press, 1975.
- "It Ain't Necessarily So" (1962b). Reprinted in *Mathematics, Matter and Method, Philosophical Papers*, Vol. 1. Cambridge, U.K.: Cambridge University Press, 1975.
- "Dreaming and 'Depth Grammar'" (1962c). In *Philosophical Papers*, Vol. 2. Cambridge, U.K.: Cambridge University Press, 1975.
- "What Theories Are Not" (1962d). In *Philosophical Papers*, Vol. 1. Cambridge, U.K.: Cambridge University Press, 1975.
- "The Meaning of 'Meaning.'" In *Philosophical Papers*, Vol. 2. Cambridge, U.K.: Cambridge University Press, 1975.
- "Realism and Reason" (1976). Reprinted in *Meaning and the Moral Sciences*. London: Routledge and Kegan Paul, 1978.
- Reason, Truth and History*. Cambridge, U.K.: Cambridge University Press, 1981.
- "Models and Reality." Reprinted in *Realism and Reason, Philosophical Papers*, Vol. 3. Cambridge, U.K.: Cambridge University Press, 1983.
- The Many Faces of Realism*. LaSalle, IL: Open Court, 1987.
- "Realism with a Human Face." In *Realism with a Human Face, Philosophical Papers*, Vol. 4. Cambridge, MA: Harvard University Press, 1990.
- Renewing Philosophy*. Cambridge, MA: Harvard University Press, 1992.



*Words and Life*. Cambridge, MA: Harvard University Press, 1994.

#### WORKS ON PUTNAM

Clark, P., and B. Hale, eds. *Reading Putnam*. Oxford: Blackwell, 1994.

Ebbs, G. *Rule-Following and Realism*. Cambridge, MA: Harvard University Press, 1997.

Rorty, R. "Solidarity or Objectivity." In *Post-analytic Philosophy*, edited by J. Rajchman and C. West. New York: Columbia University Press, 1985.

*Charles Travis (1996)*

## PYRRHO

(c. 360–c. 270 BCE)

Pyrrho of Elis is much less well known than the eponymous philosophy he inspired, Pyrrhonism. Diogenes Laertius, in his biography of Pyrrho in *Lives of Eminent Philosophers*, offers his usual mixture of anecdote, scandal, and unreliable doctrinal information (9.61–70). Thus Pyrrho was said to have had no concern for his own safety and to have been rescued from precipices and oncoming traffic by the timely interventions of his (presumably nonskeptical) friends (9.62). Aenesidemus rejected such fables, "saying that while he did philosophize in accordance with suspension of judgment, he did not act in a heedless manner" (9.62). Such stories, as well as ones presumably designed to exalt his image, such as one that claims that he demonstrated his unworldly indifference by washing pigs (9.66), are apocryphal, but they indicate what others apparently took his skeptical detachment to amount to.

He is also said to have traveled with Anaxarchus as far as India, where he consorted with the "gymnosophists," the naked philosophers, which led him "to philosophize in the noblest manner, adopting non-apprehension (*akatalēpsia*) and suspension of judgment (*epochē*); he said that nothing was good or bad, just or unjust, and that in all cases nothing is really true, but that men act by law and custom in all cases; for each thing is no more (*ou mallon*) thus than not" (9.61). This suggests that Pyrrho did indeed institute much of what was to become distinctive about later Pyrrhonism, but it also implies that he was primarily concerned with ethical questions (broadly construed)—an impression confirmed by some later evidence, most noticeably that of Cicero, who treats him exclusively as originating an obsolescent quietist ethics.

However, other testimonies suggest a broader engagement with more general epistemological themes. Pyrrho himself wrote nothing, but a disciple, Timon of Phlius, lauded his master in both prose and poetry, of which some seventy-one fragments survive. Some are devoted to exalting Pyrrho's imperturbable and noble character at the expense of the "vanity" of other philosophers. But more important for an assessment of Pyrrho's philosophical position is a report of a passage from one of Timon's prose works, embedded in an antiskeptical tract of the Aristotelian Aristocles of Messene, and itself preserved by way of the Christian Eusebius. (Such is the tortuous route of the early skeptical tradition, and attempting to purge such intrinsically hostile reports of later accretions of distortion and selectivity is a serious scholarly challenge.)

Aristocles reports Pyrrho as holding that "we are so constituted as to know nothing," and hence that all inquiry is pointless. Since for Sextus Empiricus and other later skeptics, the first claim would be unacceptably (if negatively) dogmatic, and since Sextus defines the Pyrrhonian way as one of continued (if unrequited) inquiry, this may be a misrepresentation on Aristocles' part. But it is equally possible that Pyrrho did conclude that nothing was knowable and inquiry futile, and that the subsequent rejection of these views was a later development. This latter possibility gains support from recent suggestions (see Bett 2000) that Pyrrho was not in fact the skeptical hero that later skeptics, such as Sextus, wanted to paint him as for ideological reasons of their own. Both Aenesidemus and Sextus treated Pyrrho as an archetypical role model. But Sextus, while remarking that "the skeptic school ... is called 'Pyrrhonian' from the fact that Pyrrho seems to have taken up scepticism more thoroughly and conspicuously than any of his predecessors" (1.7), rarely mentions him by name elsewhere in his large *oeuvre*. And Diogenes reports that one Numenius claimed that Pyrrho dogmatized, that is, held positive tenets (9.68). If this is right, then Pyrrho might have become a model because of his legendary imperturbability, rather than because of any practice of skeptical argumentation. Working against this view are some of the citations quoted and Diogenes' claim that Pyrrho was skilled in dialectical argument (9.64), which was to be the hallmark of later skepticisms.

At all events, Timon, as reported by Aristocles, states that according to Pyrrho, "[i] One must consider three questions: First, how are things by nature? Second, what should our attitudes toward them be? Third, what will be the result of adopting such an attitude? ... [ii] Pyrrho

declared all things equally indifferent, unmeasurable, and undecidable; [iii] for this reason [or since] neither our sensations nor our judgments are true or false. [iv] Consequently, we should not put our trust in them but should be without opinion, uncommitted, and unswayed, saying of each thing no more [*ou mallon*] [a] that it is than [b] that it is not, or [c] that it both is and is not, or [d] that it neither is nor is not. [v] For those thus disposed, the consequence will be first nonassertion [*aphasia*] then tranquility [*ataraxia*]” (Aristocles, in Eusebius, 14.18.2; = Long and Sedley Fr. 1F–5; my translation).

This is by far our most important philosophical testimony for Pyrrho, and it is based upon (and perhaps reports verbatim) a text of Pyrrho’s own pupil. It is, however, multiply difficult to interpret. The following interpretation is one way to try to make sense out of this difficult passage. The “things” (*pragmata*) of [i] may be states of affairs in the world or, more vaguely, subjects of possible cognition. “Indifferent” (*adiaphora*) in [ii] may perhaps be better rendered “undifferentiable” (thus making the claim not about the metaphysical condition of things but rather about our epistemic position with regard to *pragmata*). “Since,” the alternative connective of [iii], represents a textual conjecture that has won some scholarly support, and it has the obvious effect of reversing the direction of dependence between [ii] and [iii]. The scope of “no more” in [iv] is unclear (the sentence is syntactically ambiguous); it may govern four disjuncts (including disjuncts [c] and [d]), rather than just the first two. Finally, the precise sense of “aphasia” in [v] is disputed.

With the connective “for this reason” in [iii] (the reading of the manuscripts), the assertion is that the indeterminacy of things renders our sensations and judgments about them neither true nor false. At first sight, this seems to be a strange inference. (Does not indeterminacy simply render them false, insofar as they make positive claims that fail to correspond to the indeterminate facts?) Yet the assertion can be made intelligible if we suppose that for a claim of the form “ $x$  is  $F$ ” to be true,  $x$  must be unequivocally  $F$ , and equally for it to be false,  $x$  must be wholly not  $F$ . Thus, because of the indeterminacy in things, any unequivocal statement of the form “ $x$  is  $F$ ” will be partly true and partly false, and hence neither wholly true nor wholly false. Consequently, withholding strong belief and commitment makes sense. (In favor of the manuscript reading, the sequence of consequence is maintained: from states of affairs, via epistemic consequences, to epistemic attitude, to pragmatic consequence.) In addition, the account of truth just given

supports the interpretation that “no more” in [iv] has narrow scope, that is, that [c] and [d] are alternative ways of describing the counterpoise between [a] and [b]. To get this interpretation, we have to understand “it is” to mean “it is unequivocally” in [a] and [b], but not in [c] and [d].

The upshot, then, is that we will make no statements; not that we will literally say nothing, but that we will express no strong commitment to the unequivocal truth of our first-order remarks about the world. And when we attain this state, tranquility will follow like a shadow. As later skeptics put it, once one stops seeking to make (and support) unequivocal claims about the world, all one’s initial anxiety (apparently caused by the second-order belief that there *should* be answers to such questions and the consequent frustration of not finding them) vanishes. If all this is right, then Pyrrho really was a recognizable precursor to the later skepticism that took his name. Yet Pyrrho was not a thoroughgoing skeptic, for as the interpretation offered above suggests (but does not demand), Pyrrho did commit himself to speculations at least about the actual (Heraclitean) state of affairs of things in a way that Sextus Empiricus at least (although perhaps not Aenesidemus) would have found anathema.

**See also** Aenesidemus; Ancient Skepticism; Sextus Empiricus; Timon of Phlius.

### Bibliography

- Annas, J., and J. Barnes. *Outlines of Pyrrhonism*. Cambridge, U.K.: Cambridge University Press, 1994.
- Bett, Richard. *Pyrrho, His Antecedents, and His Legacy*. Oxford, U.K.: Oxford University Press, 2000.
- Diogenes Laertius. *Lives of Eminent Philosophers*. Translated by R. D. Hicks. Cambridge, MA: Harvard University Press, 1972.
- Eusebius of Caesarea. *Preparation for the Gospel*. Translated by Edwin Hamilton Gifford. Oxford, U.K.: Clarendon Press, 1903.
- Hankinson, R. J. *The Sceptics*. London: Routledge, 1995.

R. J. Hankinson (2005)

## PYRRHONIAN PROBLEMATIC, THE

### KNOWLEDGE AND JUSTIFICATION

If a belief is to count as knowledge, then it must be true. But truth is not enough: lucky guesses and, more generally, beliefs that are only accidentally related to the facts

they purport to describe do not amount to knowledge. What else, besides truth, is needed for a belief to count as knowledge, then? There is no agreement regarding how to fully answer that question, but there is a line of thought regarding how to begin such an answer that is widely shared: for a belief to amount to knowledge it has to be justified or supported by reasons, or rationally grounded, or warranted, or have some sort of positive epistemic status. (These, and other, words are sometimes used as synonyms, whereas sometimes they are intended to mark important epistemological distinctions. I use them interchangeably.) The justification in question here is usually qualified as epistemic, to distinguish it from the kind of justification that, for example, an assassin's mother might have in believing that her son is innocent despite mounting evidence against him.

It is possible to adopt many different attitudes with respect to any proposition  $p$  (say, the proposition that Paris is the capital of France). For instance, it is possible to believe that Paris is the capital of France, to be happy that Paris is the capital of France, to hope that Paris is the capital of France, and so on. Some of these attitudes can be called doxastic attitudes. What distinguishes a doxastic attitude from other attitudes we can adopt toward a proposition is that one can be justified or unjustified (in the epistemic sense) in adopting a doxastic attitude.

There are three basic doxastic attitudes: belief, disbelief, and suspension of judgment. (It might be that there are other attitudes that we might be justified or unjustified, in the epistemic sense, in adopting, but belief, disbelief, and suspension of judgment are basic in the sense that any other doxastic attitude will be such only because it entails one of these three basic attitudes.) To disbelieve that Paris is the capital of France is to believe that it is not true that Paris is the capital of France (and so, depending on how you count, you might think that there are only two basic doxastic attitudes: belief and suspension of judgment). To suspend judgment with respect to the proposition that Paris is the capital of France is to be in a mental state that is opposed both to believing and disbelieving the proposition. Suspension of judgment must therefore be carefully distinguished from having no attitude whatsoever with respect to a certain proposition. There is a difference between never having considered the question whether there is an even number of stars in the Milky Way and, having considered it, suspending judgment with respect to the question.

## ACADEMIC AND PYRRHONIAN SKEPTICISM

If the connection between knowledge and justification presented earlier is correct, then we can know a proposition only if we are justified in believing it. Skepticism with respect to a range of propositions is the claim that the only justified attitude with respect to the propositions in that range is to suspend judgment. We are all skeptics, in this sense, with respect to some range of propositions. For instance, it seems obvious that the only correct attitude with respect to the proposition that there is an even number of stars in the Milky Way, once we have considered it, is to suspend judgment. This is ordinary skepticism. But most of us are nonskeptics with respect to many propositions. For instance, it seems obvious that the only justified attitude with respect to the proposition that Paris is the capital of France is to believe it, whereas the only justified attitude with respect to the proposition that Tony Blair is the president of the United States is to disbelieve it. Philosophical skepticism extends well beyond ordinary skepticism, claiming that we should suspend judgment with respect to propositions that we ordinarily think we are justified in believing.

It is customary to distinguish between two different kinds of philosophical skepticism, which can be called, following an ancient tradition, Academic skepticism and Pyrrhonian skepticism. Academic skepticism referred originally to a phase in the history of Plato's Academy that stretched approximately from the third to the early first century BCE. The main figures of Academic skepticism were Arcesilaus (mid-third century BCE), Carneades (mid-second century BCE), and Clitomachus (d. 110/109 BCE). The main sources for Pyrrhonian skepticism are the writings of Sextus Empiricus in the late second century CE.

Academic and Pyrrhonian skepticism differ in the scope of propositions that, according to them, we should suspend judgment about. Let's call those propositions that do not contain any epistemic concepts ordinary propositions and let's call those propositions to the effect that someone knows an ordinary proposition can be called epistemic propositions.

Academic skeptics think that the only justified attitude with respect to most (perhaps all) ordinary propositions is suspension of judgment. However, Academic skeptics do not suspend judgment with respect to epistemic propositions: On the contrary, they think that the only justified attitude with respect to them is to disbelieve them—that is, they think that we are justified in believing that we do not know almost anything of what we take

ourselves to know. (When contemporary authors discuss skepticism, chances are they are referring to this aspect of Academic skepticism: to the claim that we do not know certain propositions that we ordinarily take ourselves to know. However, the tradition is to classify as a skeptic with respect to a certain proposition only someone who thinks we should suspend judgment with respect to that proposition, not someone who thinks that we should dissent from it.)

Pyrrhonian skeptics, meanwhile, extend their skepticism to epistemic propositions as well. Both Academic and Pyrrhonian skeptics leave it open whether Paris is the capital of France or not: maybe it is, maybe it is not, but we are not justified in believing that it is or believing that it is not. According to Pyrrhonian skeptics it is also an open question whether we know that Paris is the capital of France: maybe we do, maybe we do not, but we are not justified in believing that we do or that we do not. Academic skeptics, on the contrary, do not leave this question open: they think we are justified in believing that we do not know that Paris is the capital of France.

### THE MODES OF AGRIPPA

From now on, the focus will be on Pyrrhonian skepticism exclusively. The Pyrrhonians had a number of ways, or modes, to induce suspension of judgment. The importance of Pyrrhonian skepticism to contemporary epistemology derives primarily from these modes, and in particular from a subset of them referred to collectively as the modes of Agrippa. There are five modes associated with Agrippa, but three of them are the most important: the mode of hypothesis (or unsupported assertion), the mode of circularity (reciprocal), and the mode of regression to infinity.

The three modes of Agrippa function together in the following way. Whenever the dogmatist (Sextus refers to those who are not skeptics as dogmatists) asserts his or her belief in a proposition  $p_1$ , the Pyrrhonian will challenge that assertion, asking the dogmatist to justify  $p_1$ , to give reasons for thinking that it is true. The dogmatist will then either decline to answer the challenge or adduce another proposition  $p_2$  in support of  $p_1$ . If the dogmatist refuses to answer the challenge, the Pyrrhonian will be satisfied that the only justified attitude to take with respect to  $p_1$  is to suspend judgment, because no reason for it has been given (thus appealing to the mode of hypothesis). If the dogmatist adduces another proposition  $p_2$  in support of  $p_1$ , then either  $p_2$  will be identical to  $p_1$  or it will be a different proposition. If  $p_2$  is the same proposition as  $p_1$ , then the Pyrrhonian will also suspend

judgment with respect to  $p_1$ , because no proposition can support itself (thus appealing to the mode of circularity). If, however,  $p_2$  is different from  $p_1$ , then the Pyrrhonian will ask the dogmatist to justify his or her assertion of  $p_2$ . And now the dogmatist offers no reason in support of  $p_2$ , offers  $p_2$  itself or  $p_1$  as a reason, or adduces yet another proposition  $p_3$ , different from both  $p_1$  and  $p_2$ . If the dogmatist offers no reason for  $p_2$ , then the Pyrrhonian will invoke the mode of hypothesis again and suspend judgment in accordance with it; if either  $p_2$  itself or  $p_1$  is offered as a reason to believe in  $p_1$ , then the Pyrrhonian will invoke the mode of circularity and suspend judgment in accordance with it (because not only can no proposition be a reason for believing in itself but also no genuine chain of reasons can loop); and, finally, if the dogmatist offers yet another proposition  $p_3$ , different from both  $p_1$  and  $p_2$ , as a reason to believe  $p_2$ , then the same three possibilities that arose with respect to  $p_2$  will arise with respect to  $p_3$ .

The dogmatist will not be able to continue offering different propositions in response to the Pyrrhonian challenge forever—eventually, either no reason will be offered, or a proposition that has already made an appearance will be mentioned again. The Pyrrhonian refers to this impossibility of actually offering a different proposition each time a reason is needed as the mode of infinite regression. The three Pyrrhonian modes, then, work in tandem to induce suspension of judgment with respect to any proposition whatsoever.

### AGRIPPA'S TRILEMMA

The Pyrrhonian use of the three modes of Agrippa to induce suspension of judgment can be presented in the form of an argument, called Agrippa's trilemma. It is at least somewhat misleading to present the Pyrrhonian position in terms of an argument, because in presenting an argument one is usually committed to the truth of its premises and conclusion, whereas Pyrrhonian skeptics would suspend judgment with respect to them. Nevertheless, presenting the Pyrrhonian problematic in the form of an argument does not do much violence to this skeptical position, because what is important is not whether the Pyrrhonian skeptics themselves accept the premises or the validity of the argument, but whether their audience does. Problems still remain regarding the coherence of anyone (be they Pyrrhonian skeptics or not) who accepts the soundness of an argument whose conclusion is that we are not justified in believing anything. It is doubtful, though, whether anyone accepts Agrippa's trilemma: "Dogmatists" certainly do not, and neither do Pyrrhon-

ian skeptics. It is not a coincidence that Wittgenstein's dictum about throwing the ladder after using it to climb echoes Sextus's less-pleasing image of the laxative that purges itself together with the "humours" of the body it is designed to expel. Still, even if we do not think that the argument is sound, we stand to learn something interesting about the structure of an epistemological theory—because each of the premises of the apparently valid argument looks plausible at first sight.

Before presenting a reconstruction of Agrippa's trilemma some definitions need to be introduced. Say that a belief is inferentially justified if and only if it is justified (at least in part) in virtue of its relations to other beliefs. A justified basic belief, by contrast, is a belief that is justified but not in virtue of its relations to other beliefs. An inferential chain is a set of beliefs such that every member of the set is allegedly related to at least one other member by the relation is justified by. Agrippa's trilemma, then, can be presented thus:

- (1) If a belief is justified, then it is either a basic justified belief or an inferentially justified belief.
- (2) There are no basic justified beliefs.

Therefore,

- (3) If a belief is justified, then it is justified in virtue of belonging to an inferential chain.
- (4) All inferential chains are such that either (a) they contain an infinite number of beliefs; (b) they contain circles; or (c) they contain beliefs that are not justified.
- (5) No belief is justified in virtue of belonging to an infinite inferential chain.
- (6) No belief is justified in virtue of belonging to a circular inferential chain.
- (7) No belief is justified in virtue of belonging to an inferential chain that contains unjustified beliefs.

Therefore,

- (8) There are no justified beliefs.

Premise (1) is beyond reproach, given our previous definitions. Premise (2) is justified by the mode of hypothesis. Step (3) of the argument follows from (1) and (2). Premise (4) is also beyond reproach—the only remaining possible structure for an inferential chain to have is to contain basic justified beliefs, but there are none of those according to premise (2). Premise (5) is justified by appeal to the mode of infinite regression, and (6) is justified by appeal to the mode of circularity.

Premise (7) might seem to be a truism, but some authors have argued that denying it is the only plausible way out of Pyrrhonian skepticism.

It is interesting to note that Agrippa's trilemma is perfectly general; in particular, it applies to philosophical positions as well as to ordinary propositions. In fact, when Agrippa's trilemma is applied to epistemological theories themselves, the result is called "the problem of the criterion."

Many contemporary epistemological positions can be stated as a reaction to Agrippa's trilemma. In fact, all of premises (2), (5), (6), and (7) have been rejected by different philosophers at one time or another. In the remainder of this entry, we examine each of these responses.

### REJECTING PREMISE (2): FOUNDATIONALISM

Foundationalists claim that there are basic justified beliefs—beliefs that are justified but not in virtue of their relations to other beliefs. In fact, according to foundationalists all justified beliefs are either basic beliefs or are justified in virtue of being inferentially related to a justified belief (or to some justified beliefs). This is where foundationalism gets its name: The edifice of justified beliefs has its foundation in basic beliefs.

But how do foundationalists respond to the mode of hypothesis? If a belief is not justified by another belief, then is it not just a blind assertion? If basic beliefs are justified but not by other beliefs, then how are they justified? What else besides beliefs is there that can justify beliefs?

To this last question, many foundationalists reply: experience (if they are talking about empirical knowledge, of course; *a priori* knowledge raises interesting problems of its own, and it is also subject to Agrippa's trilemma). To a rough first approximation that glosses over many important philosophical issues, experiences are mental states that, like beliefs, aim to represent the world as it is, and, like beliefs, can fail in achieving that aim—that is, experiences can misrepresent. Nevertheless, experiences are not to be identified with beliefs, for it is possible to have an experience as of, for example, facing two lines that differ in length without having the belief that one is facing two lines that differ in length—a combination of mental states that anyone familiar with the Müller-Lyer illusion will recognize.

There are three important questions that any foundationalist has to answer. First, what kinds of beliefs do experiences justify? Second, how must inferentially acquired beliefs be related to basic beliefs for them to be

justified? Third, in virtue of what do experiences justify beliefs?

**TRADITIONAL AND MODERATE FOUNDATIONALISM.** With respect to the first question, we can distinguish between traditional foundationalism and moderate foundationalism. Traditional foundationalists think that basic beliefs are beliefs about experiences, whereas moderate foundationalists think that experience can justify beliefs about the external world. Take, for example, the experience that you typically have when looking at a tomato under good perceptual conditions—an experience that, remember, can be had even if no tomato is actually there. A moderate foundationalist would say that that experience justifies you in believing that there is a tomato in front of you. The traditional foundationalist, however, would say that the experience justifies you only in believing that you have an experience as of a tomato in front of you. You may well be justified in believing that there is a tomato in front of you, but only inferentially.

A traditional argument in favor of traditional foundationalism relies on the fact that whereas you can be mistaken regarding whether there is a tomato in front of you when you have an experience as of facing a tomato, you cannot, in the same situation, be mistaken regarding whether you are undergoing such an experience. From the point of view of traditional foundationalism, this fact indicates that the moderate foundationalist is taking an unnecessary epistemic risk—the risk of having a foundation composed of false beliefs.

The moderate foundationalist can reply that the traditional foundationalist must undertake a similar risk. For, while it is true that if one is undergoing a certain experience then one cannot be mistaken in thinking that one is undergoing that experience, one can still be mistaken about one's experiences—for instance, perhaps one can believe that one is in pain even if the experience that one is undergoing is actually one of feeling acutely uncomfortable. And if it were just as difficult to distinguish between the true and the false in the realm of beliefs about our own experiences as it is in the realm of beliefs about the external world, then we could be wrong about which of our own beliefs are basically justified and which are not. If this kind of metafallibilism is accepted, then why not accept the further kind according to which basic justified beliefs can be false? Of course, the resolution of this dispute depends on whether, as the moderate believes, we can be mistaken about our own experiences.

**DEDUCTIVIST AND NONDEDUCTIVIST FOUNDATIONALISM.** What about our second question: How must basic beliefs be related to inferentially justified beliefs? Here, too, there are two different kinds of foundationalism: deductivism and nondeductivism. According to the deductivist the only way in which a (possibly one-membered) set of basic justified beliefs can justify another belief is by logically entailing that other belief. In other words, there has to be a valid argument whose premises are all basic justified beliefs and whose conclusion is the inferentially justified belief in question. Given that the argument is valid, the truth of the premises guarantees the truth of the conclusion—it is impossible for all the premises to be true while the conclusion is false. Nondeductivism allows relations other than logical entailment as possible justificatory relations. For instance, many foundationalists will claim that good inductive inferences from basic justified beliefs provide their conclusions with justification—even though inductive arguments are not valid, that is, even though it is possible for all the premises of a good inductive argument to be true while its conclusion is false. Although these are independent distinctions, traditional foundationalists tend to be deductivists, whereas moderate foundationalists tend to be nondeductivists. Notice that for a traditional, deductivist foundationalist, there cannot be false justified beliefs. Many contemporary epistemologists would shy away from this strong form of infallibilism and take that consequence to be an argument against the conjunction of traditional foundationalism and deductivism.

**PRIMITIVIST, INTERNALIST, AND EXTERNALIST FOUNDATIONALISM.** The question that is most interesting from the point of view of the Pyrrhonian problematic is our third one: What is it about the relation between an experience and a belief that, according to the foundationalist, allows the former to justify the latter? (Analogous questions apply to nonfoundationalist positions too, and the discussion to follow is not restricted to the specific case of foundationalism.) There are three different proposals about how to answer this question that are the most prominent. The principles that assert that a subject is justified in having a certain belief given that he or she is undergoing a certain experience can be called epistemic principles. Our third question can then be stated as follows: What makes epistemic principles true?

The first proposal, which we shall call primitivism, claims that the question cannot have an intelligible answer. There is no more basic fact in virtue of which epistemic principles obtain. They describe bedrock facts, not to be explained in terms of anything else, but are

instead to be used to explain other facts. Epistemological theorizing, according to the primitivist, ends with the discovery of the correct epistemic principles.

The other two positions are nonprimitivist. Internalist nonprimitivism holds that epistemic principles are true in virtue of facts about ourselves—for instance, one prominent internalist view is that which epistemic principles are true for a given subject is determined by which epistemic principles that subject would accept under deep reflection. Externalist nonprimitivism holds that epistemic principles are true in virtue of facts that are not about ourselves—for instance, one prominent externalist view is that certain experiences provide justification for certain beliefs because the obtaining of those experiences is reliably connected to the truth of those beliefs (reliabilism), or because i.e., it could not easily happen that those experiences obtain without those beliefs being true (an appeal to “sensitivity” or “safety” conditionals).

Both externalists and internalists think that primitivists are overlooking real facts, whereas primitivists think that there are fewer things in heaven and earth than are dreamt of in nonprimitivist philosophy. Within the nonprimitivist camp externalists think that internalists have too subjective a conception of epistemology—to some extent, thinking it so, or being disposed to think it so under conditions of deep reflection, makes it so for at least some traditional internalists. Internalists, for their part, are likely to think that externalists are no longer engaged in the same project that both skeptics and internalist epistemologists are engaged in, the project of determining “from the inside” whether one’s beliefs are justified or amount to knowledge, because the obtaining of a relation between a subject’s belief and the external world is something that the subject is in no position to ascertain “from the inside.”

#### REJECTING PREMISE (5): INFINITISM

Infinetism, the claim that infinite evidential chains can provide justification to their members, is the answer to Agrippa’s trilemma that has received the least attention in the literature. This is due, at least in part, to the fact that infinitism has to deal with what might seem like formidable obstacles. For instance, it seems that no one actually has an infinite number of beliefs. To this objection, the infinitist is likely to reply that actually occurring beliefs are not needed, only implicit beliefs that are available to the subject to continue constructing his or her inferential chain if called on to do so (by others or by him- or herself). The plausibility of this reply depends on whether good sense can be made of the notion of implicit belief

and the notion of an implicit belief’s being available for a subject.

Even leaving that problem aside, the infinitist, like the coherentist, maintains that justification can arise merely in virtue of relations among beliefs. Infinitists will then have to respond to many of the same objections that are leveled against coherentism—in particular, they would have to respond to the isolation objection mentioned in the next section.

#### REJECTING PREMISE (6): COHERENTISM

Coherentists reject two related features of the picture of evidential reasons that underlies Agrippa’s trilemma. The first feature is the idea that justification is an asymmetrical relation: if a belief  $p_1$  justifies a different belief  $p_2$ , then  $p_2$  does not justify  $p_1$ . The second feature is the idea that the unit of justification is the individual belief. Putting these two rejections together, the coherentist believes that justification is a symmetrical and holistic matter. It is not individual beliefs that are justified in the primary sense of the word, but only complete systems of beliefs—individual beliefs are justified, when they are, in virtue of belonging to a justified system of beliefs. The central coherentist notion of justification is best taken to be a comparative one: A system of beliefs B1 is better justified than a system of beliefs B2 if and only if B1 has a greater degree of internal coherence than B2. One crucial question that coherentists have to answer, of course, is what it takes for one system of beliefs to have a greater degree of coherence than another. Many coherentists have thought that explanatory relations will be crucial in elucidating the notion of coherence: The more explanatorily integrated a system is, the more coherence it displays.

The main objection that coherentists have to answer is called the isolation objection. The objection centers on the fact that, according to the coherentist, the justification of a system of beliefs is entirely a matter of relations among the beliefs constituting the system. But this runs against the strong intuition that experience has an important role to play in the justification of beliefs. To illustrate the problem, suppose that you and I both have a highly coherent set of beliefs—your system, it is safe to assume, contains the belief that you are reading, whereas mine does not, and it contains instead the belief that I am swimming (because, let us suppose, I am swimming right now). Suppose now that we switch systems of beliefs—somehow, you come to have my set of beliefs and I come to have yours. Given that coherence is entirely a matter of relations among beliefs, your system will be as coherent in

my mind as it was in yours, and vice versa. And yet, our beliefs are now completely unjustified—there you are, reading, believing that you are swimming, and here I am, swimming, believing that I am reading. In other words, certain transformations that preserve coherence in a system of beliefs do not seem to preserve justification.

In reply, coherentists argue that it is possible to give experience a role without sacrificing the idea that justification is entirely a matter of relations among beliefs—one idea is to require that any minimally acceptable system of beliefs contain beliefs about the experiences that the subject is undergoing. It is fair to say that there is no agreement regarding whether this move can solve the problem.

### REJECTING PREMISE (7): POSITISM

One position that can be traced back to some ideas in Ludwig Josef Johann Wittgenstein's *On Certainty* (published posthumously in 1969)—and, perhaps, also to José Ortega y Gasset's *Ideas y Creencias* (1940)—is that evidential chains have to terminate in beliefs that are not properly said to be either justified or unjustified. This position, which we shall call positism (not to be confused with positivism), shares many features with foundationalism: for instance, both positists and foundationalists agree that inferential chains have to be finite and noncircular. But, whereas the foundationalist thinks that the starting points of inferential chains are beliefs that are justified by something other than beliefs, the positist thinks that the starting points of inferential chains are beliefs that are not justified by anything—they are posits that we have to believe without justification. Despite this difference between the positist and the foundationalist, the positions are structurally similar enough that analogues of the questions posed to the foundationalist can be asked of the positist.

First, then, which beliefs are such that they are not justified and yet are the starting points of every inferential chain—in other words, how do we identify which are the posits? One answer that can be gleaned from Wittgenstein's *On Certainty*, which we will call relativistic positism, is that this is a matter that is relative both to time and society, because what the posits are is determined by some function of the actual positing practices of the members of one's society at a certain time. Thus, according to Wittgenstein the proposition that no one has been to the moon was a posit for a certain long period of time—it was a proposition that no one felt the need to justify, and that was presupposed in many justificatory practices. For obvious reasons, though, that proposition

can no longer appropriately function as a posit. Other epistemologists, nonrelativistic positists, think that which beliefs are properly posited depends on some objective truth about which beliefs have to be presupposed to engage in the practice of justifying beliefs at all. One prime candidate for playing this role is the first-person belief that I am not being deceived by an evil demon into thinking that I am a normally embodied and situated human being.

The second question, regarding how posits must be related to inferred beliefs to justify them, can receive answers that are completely analogous to the foundationalists'. The third question, applied to positism, is the question why certain beliefs are properly posited. Relativistic positists answer that this is so because of a certain societal fact: because they are taken to be so by an appropriate subsector of a certain society at a certain time. Nonrelativistic positists answer that a certain belief is properly taken as a posit just in case every justificatory act that we engage in presupposes that the belief in question is true.

One objection that positists of both sorts have to face is that they are transforming a doxastic necessity into an epistemic virtue—that is, they are concluding that certain beliefs can properly serve as the starting points of inferential chains because that is how in fact they are treated (relativistic positism) or because otherwise it would not be possible to engage in inferential practices at all (nonrelativistic positism). The Pyrrhonian skeptic, of course, will reply that the mere fact that most members of a society accept a certain belief without justification, or even the fact that if we do not do so then we cannot justify anything else, does not mean that it should be accepted without justification.

### CONCLUSION

Perhaps one of the most interesting developments in relation to the Pyrrhonian problematic is that more and more epistemologists are arguing that the proper way to reply to Agrippa's trilemma is to combine some of the positions that, for ease of exposition, we have presented as mutually exclusive (this development is explicit in contemporary authors such as Sosa, but, some will argue, it is already present in Descartes). Thus, for example, many contemporary epistemologists put forward theories that contain elements of both internalism and externalism, or foundationalism and coherentism. It is a testament to the endurance of the Pyrrhonian problematic that philosophers continue in this way to grapple with it.



*See also* Agrippa; Ancient Skepticism; Arcesilaus; Carneades; Classical Foundationalism; Coherentism; Descartes, René; Greek Academy; Internalism versus Externalism; Ortega y Gasset, José; Plato; Pyrrho; Sextus Empiricus; Skepticism, History of; Wittgenstein, Ludwig Josef Johann.

### **Bibliography**

- Barnes, Jonathan. *The Toils of Scepticism*. New York: Cambridge University Press, 1990.
- BonJour, Lawrence. *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press, 1985.
- Burnyeat, Myles. "Can the Skeptic Live His Skepticism?" In *Doubt and Dogmatism: Studies in Hellenistic Epistemology*, edited by Malcolm Schofield, Myles Burnyeat, and Jonathan Barnes. Oxford, U.K.: Clarendon Press, 1980.
- DePaul, Michael R., ed. *Resurrecting Old-Fashioned Foundationalism*. Lanham, MD: Rowman and Littlefield, 2001.
- Fogelin, Robert J. *Pyrrhonian Reflections on Knowledge and Justification*. New York: Oxford University Press, 1994.
- Klein, Peter. "Human Knowledge and the Infinite Regress of Reasons." *Philosophical Perspectives* 13 (1999): 297–325.
- Ortega y Gasset, José. *Ideas y Creencias*. Buenos Aires: Espasa Calpe, 1940.
- Sextus Empiricus. *Outlines of Scepticism*. 2nd ed., edited by Julia Annas and Jonathan Barnes. New York: Cambridge University Press, 2000.
- Sosa, Ernest. "How to Resolve the Pyrrhonian Problematic: A Lesson from Descartes." *Philosophical Studies* 85 (2–3) (1997): 229–249.
- Wittgenstein, Ludwig. *On Certainty (Über Gewissheit)*, edited by G. E. M. Anscombe and G. H. von Wright. Translated by Denis Paul and G. E. M. Anscombe. Oxford: Basil Blackwell, 1969–1975.

*Juan Comesaña (2005)*

## PYTHAGORAS AND PYTHAGOREANISM

Pythagoras was an Ionian Greek born on the island of Samos, probably about 570 BCE. His dislike of the policies of the Samian tyrant Polycrates caused him to immigrate to Crotona in southern Italy. There he founded a society with religious and political, as well as philosophical, aims that gained power in the city and considerably extended its influence over the surrounding area. A certain Cylon, however, stirred up a revolt against the society in which a number of its leading members were killed, and Pythagoras retired to Metapontum. The community recovered its influence until a more serious persecution took place in the middle of the fifth century, from which the survivors scattered to various parts of the Greek

world—notably Thebes, Phleius, and Tarentum. In these places "they preserved their original ways and their science, although the sect was dwindling, until, not ignobly, they died out" (in the late fourth century), to quote the epitaph written by a contemporary.

### NATURE OF THE EVIDENCE

The obstacles to an appraisal of classical Pythagoreanism are formidable. There exists no Pythagorean literature before Plato, and it was said that little had been written, owing to a rule of secrecy. Information from the Christian era is abundant but highly suspect. Pythagoras himself, though a fully historical figure, underwent a kind of canonization. His life was quickly obscured by legend, and piety attributed all the school's teaching to him personally. Moreover, the dispersion of the school inevitably led to divergences of doctrine in the various groups. Aristotle makes it clear that by the late fifth century some Pythagoreans were teaching one thing and some another. A further reason for division was that the universal genius of Pythagoras, for whom religion and science were two aspects of the same integrated worldview, was beyond the scope of lesser men. Some naturally inclined more to the religious and superstitious; others, to the intellectual and scientific side, as is confirmed by later references to the division between *acusmatici* and *mathematici*.

As early evidence there are several references to Pythagoras in works of his contemporaries or near contemporaries (for instance, Xenophanes satirized his belief in the transmigration of souls), a valuable reference in Plato to the relationship between astronomy and harmonics in the Pythagorean system, a quantity of information from Aristotle (who at least would not confuse the Pythagoreans with Plato, as later writers excusably did), and some quotations from pupils of Aristotle who were personally acquainted with the last generation of the school.

Given the nature of the sources, the following is a fairly conservative summary of Pythagoreanism before Plato.

### MAN AND THE COSMOS

In contrast with the Milesians, the Pythagoreans were not motivated by disinterested scientific curiosity. For Pythagoras, philosophy was the basis of a way of life, leading to salvation of the soul. "Their whole life," said a fourth-century writer, "is ordered with a view to following God, and it is the governing principle of their philosophy." At philosophy's center, therefore, were man and his relation to other forms of life and to the cosmos. Purity

was to be sought by silence, self-examination, abstention from flesh and beans, and the observance of other primitive taboos that the Pythagoreans interpreted symbolically. Of the recognized gods they worshiped Apollo, guardian of the typically Greek ideal of moderation (“nothing too much”), of whom Pythagoras was believed to be an incarnation.

Behind both the superstition and the science was the notion of kinship or sympathy. The kinship and essential unity of all life made possible the belief in the transmigration of souls and accounted for the prohibition of meat: a sheep might house the soul of an ancestor. Not only animate nature in our sense but the whole world was akin, for the cosmos itself was a living, breathing creature. The cosmos was one, eternal, and divine; men were divided and mortal. But the essential part of man, his soul, was not mortal; it was a fragment of the divine, universal soul that was cut off and imprisoned in a mortal body. Men should therefore cultivate and purify the soul, preparing it for a return to the universal soul of which it was a part. Until then, since it was still contaminated by the body, it must tread the wheel of reincarnation, entering a new body of man or animal after the death of its previous tenement.

These tenets were also taught by the religious movement known as Orphism, from which the religious side of Pythagoreanism can hardly be separated. (Pythagoras himself was said in the fifth century to have written books under the name of Orpheus.) But whereas the Orphics sought salvation by purely religious means—sacramental ceremonies and the observance of ritual prohibitions—Pythagoras added a new way, the way of philosophy.

Philosophy, for Pythagoras as for others, meant the use of reason and observation to gain understanding of the universe. The link between this procedure and his overriding aim of salvation seems to have been the principle that like is known by like, a widespread tenet of pre-Socratic thought, common to such diverse systems as the philosophicoreligious synthesis of Empedocles and the scientific atomism of Democritus. Hence, an understanding of the divine universe would bring man’s nature closer to its own. In this conception we meet the typically Pythagorean conception of *kosmos*, a word that combines in an untranslatable way the notion of orderly arrangement or structural perfection with that of beauty. Closely linked with it is *peras*, meaning limit. An organic whole, particularly one that, like the universe, lives forever, must of necessity exhibit limit and order in the highest degree. What is unlimited has no *telos* (end) and is *a-teles*, which means both “endless” and “incomplete.” But the world is

a perfect whole, a model of order and regularity, supremely exemplified in Greek eyes by the ceaseless wheeling of the heavenly bodies in (as they believed) perfect circles, bringing about the unvarying succession of day and night and seasons. It was said of Pythagoras that he was the first to call the world *kosmos*, “from its inherent order.” By studying this order, we reproduce it in our own souls, and philosophy becomes an assimilation to the divine, as far as that is possible within the limitations imposed by our mortal bodies.

## THE DOCTRINE THAT THINGS ARE NUMBERS

The Pythagoreans studied mathematics in a cosmic context, and for them numbers always retained a mystical significance as the key to the divine cosmos. “They supposed the whole heaven to be a *harmonia* and a number,” said Aristotle. *Harmonia*, though specially applied to music, could signify any well-organized structure of parts fitted together in due proportion. Its effect in music seems to have burst on Pythagoras as a revelation of the whole cosmic system. We may accept the many later statements that he discovered the numerical ratios underlying the intervals that the Greeks called consonant and used as the basis of their scale. They involve only the numbers 1 to 4—1:2, octave; 3:2, fifth; 4:3, fourth. These numbers add up to 10, a sacred number for the Pythagoreans, which was symbolized by the dotted triangle (*tetractys*), “source and root of everlasting nature.” From the discovery that the sounds they recognized as beautiful depended on inherent, objective, mathematical order, they leaped to the conclusion that number was the key to the element of order in nature as a whole.

With this innovation the Pythagoreans would seem to have taken the momentous step from explanation in terms of matter (as the Milesians had sought it) to explanation in terms of form. Yet philosophy was not quite ready for that step, nor could the distinction between matter and form be clearly grasped. They saw simply the ultimate, single nature (*physis*) of things in their mathematical structure. There seems little doubt that probably until well on in the fifth century they thought it possible to speak of things as actually made up of “numbers” that were regarded simultaneously as units, geometrical points, and physical atoms. Lines are made of points; surfaces, of lines; solids, of surfaces; and physical bodies, of solids. In this scheme two points made a line; three, the minimum surface (triangle); four, the minimum solid (tetrahedron). A later theory spoke of the “fluxion” of point into line, line into surface, and so on, which gave a

geometrical progression (1, 2, 4, 8) instead of the arithmetical (1, 2, 3, 4), and the sequence of point, line, square, cube. Based on continuity, it seems designed to avoid the problem of incommensurable magnitudes or irrational numbers.

Whenever they were discovered (probably not much later than 450), incommensurables had dealt a blow to the original “things are numbers” doctrine, the idea that geometrical figures—and thus ultimately the physical world—are based on a series of integers. No ratio between integers can either describe the relation between the diagonal of a square and its side or serve as the basis of construction of a right triangle. If, however, magnitudes are regarded as continuous and hence infinitely divisible, the existence of incommensurable or irrational magnitudes (those which cannot be expressed as a ratio of natural numbers) could be explained and the difficulty overcome.

### THE ULTIMATE PRINCIPLES

The analysis went further than that outlined above, for numbers themselves have their elements. The ultimate principles were limit and the unlimited, which were equated with good and bad respectively; moral concepts went side by side with physical concepts in this extraordinary system. Abstractions as well as physical phenomena were equated with numbers; for instance, justice was 4, the first square number, symbolizing equality or requital. After limit and the unlimited came odd and even instances, respectively, of these two. They generated the unit (considered to be outside the number series, and both odd and even), from the unit sprang numbers, and from numbers came the world. There seems no doubt that the scheme goes back to an ultimate duality that corresponds to the moral dualism of Pythagoreanism, but one can also see how monistically minded Neoplatonic commentators could speak of the cosmos as originating from the One. In general terms, *kosmos* was achieved by the imposition of limit on the unlimited in order to make the limited, just as the imposition of definite ratios on the indefinite range of musical pitch produced the *harmonia* of the scale.

### COSMOGONY AND COSMOLOGY

Cosmogony starts with the planting of a unit in the infinite. Aristotle called it, among other things, a seed; and since limit was associated with male and unlimited with female, the Pythagoreans probably thought of the generation of the living cosmos as taking place as did that of other animals. It grows by drawing in and assimilating the

unlimited outside, that is, by conforming it to limit and giving it numerical structure. Physically the process resembles inspiration, and the unlimited is also called breath.

The unit seed had the nature of fire and in the completed cosmos (which evidently grew from the center outward) became a fire at its center. There are traces of two different cosmological schemes, a geocentric one that spoke of a fire at the center of Earth, and a more remarkable one attributed, in later sources at least, to the fifth-century Pythagorean Philolaus, which made Earth a planet. (Nicolas Copernicus in *De Revolutionibus* says that reading of this Pythagorean doctrine gave him courage to consider explaining the heavenly motions on the basis of a moving Earth.) According to this latter scheme, Earth, planets, sun, and moon—and an extra body called the counterearth—all revolved about the center of the universe, which was occupied by a fire invisible to man because he lived on the opposite side of Earth. It was known that the moon’s light is borrowed, and the idea was extended to the sun, whose heat and light were said to be reflected from the central fire. The moon was eclipsed by the interposition of both Earth and the counterearth and, according to some, of further, otherwise unknown, planetary bodies. These caused the comparatively frequent lunar eclipses.

In this system, the mixture of religion and science in Pythagoreanism is well brought out. Fire was given the central position, not for any scientific reason but because it was regarded with religious awe—and the center is the most “honorable” place. It was lauded with such titles as Hearth of the Universe, Tower of Zeus, and Throne of Zeus. Yet the same thinkers were aware that with Earth in orbit “the phenomena would not be the same” as in a geocentric scheme (presumably they were thinking of the lack of stellar parallax and variations in the apparent size of the sun and moon). They pointed out that even with a central Earth, an observer would be separated from the center by the distance of its radius, and they argued that the visible effect would be as negligible in one case as in the other. This assumes that the heavenly bodies are at vast distances from Earth; and it is not known how, if at all, this system was related to the theory later known as the harmony of the spheres.

In any case, there are many divergent versions of this doctrine. In outline, the idea was that large bodies in motion must inevitably produce a sound; that the speeds of the heavenly bodies, judged by their distances, are in the ratios of the musical consonances; and that therefore the sound made by their simultaneous revolution is con-

cordant. We do not hear it because it has been with us from birth, and sound is perceptible only by contrast with silence. It has been plausibly argued that in the original version Pythagoras, like Anaximander, took only three orbits into account (sun, moon, and all the stars); this would relate it to his original musical discovery about the fourth, fifth, and octave. Later versions speak of seven, eight (Plato), and ten orbits. In any form, the doctrine emphasizes the universal importance, in Pythagorean eyes, of mathematical and musical laws and their intimate relation to astronomy.

### NEO-PYTHAGOREANISM

The influence of Pythagorean thought on the history of philosophy and religion has been exercised largely through the medium of Plato, who enthusiastically adopted its main doctrines of the immortality of the soul, philosophy as an assimilation to the divine, and the mathematical basis of the cosmos. Later antiquity regarded him as a Pythagorean source, so that post-Platonic writings are of little help in distinguishing Pythagorean from original Platonic material in the dialogues. The Neo-Pythagorean movement, which started in the first century BCE, was an amalgam of early Pythagorean material with the teachings of Plato, the Peripatetics, and the Stoics. All of this material was credited to Pythagoras, who was revered as the revealer of esoteric religious truths. The interests of Neo-Pythagoreanism were religious and, in accordance with the prevailing tendencies of the time, it emphasized the mystical and superstitious sides of the earlier doctrine, its astral theology and number-mysticism, to the detriment of philosophical thinking. It cannot be called a system, but rather is a trend that in different forms continued until the rise of Neoplatonism in the third century CE, when it lost its identity in that broader and more powerful current. Besides contributing to Neoplatonism, it influenced Jewish thought through Philo of Alexandria and Christian thought through Clement of Alexandria. Prominent Neo-Pythagoreans were Cicero's acquaintance, Nigidius Figulus, and Apollonius of Tyana, a wandering mystic and ascetic of the first century CE, credited with miraculous and prophetic powers. Numenius of Apamea in the late second century was called both Pythagorean and Platonist, and was the immediate precursor of Neoplatonism.

**See also** Apeiron/Peras; Aristotle; Atomism; Cicero, Marcus Tullius; Clement of Alexandria; Continuity; Copernicus, Nicolas; Cosmology; Empedocles; Geometry; Leucippus and Democritus; Neoplatonism; Numenius of Apamea; Philo Judaeus; Philolaus of Croton; Plato;

Platonism and the Platonic Tradition; Pre-Socratic Philosophy; Reason; Xenophanes of Colophon.

### Bibliography

- Delatte, A. *Études sur la littérature pythagoricienne*. Paris: E. Champion, 1915. For specialists.
- Fritz, K. von. *Pythagorean Politics in South Italy: An Analysis of the Sources*. New York, 1940.
- Guthrie, W. K. C. "Pythagoras and the Pythagoreans." In his *History of Greek Philosophy*. Vol. I, 146–340. Cambridge, U.K.: Cambridge University Press, 1962. References to much of the literature of the subject will be found in this general account.
- Minar, E. L., Jr. *Early Pythagorean Politics*. Baltimore: Waverly Press, 1942.
- Morrison, J. S. "Pythagoras of Samos." *Classical Quarterly*, n.s., 6 (1956): 135–156.
- Thesleff, H. *An Introduction to the Pythagorean Writings of the Hellenistic Period*. Turku, Finland, 1961. For specialists.
- Timpanaro-Cardini, Maria. *Pitagorici: testimonianze e frammenti*. 2 vols. Florence, 1958 and 1962. Texts from *Fragmente der Vorsokratiker*, edited by H. Diels and W. Kranz, translated into Italian, with introduction and commentary.
- Van der Waerden, B. L. "Die Arithmetik der Pythagoreer." *Mathematische Annalen* 120 (1948): 127–153 and 676–700. For specialists.
- Van der Waerden, B. L. *Die Astronomie der Pythagoreer*. Amsterdam: North-Holland, 1951. For specialists. Both this and the other van der Waerden work do not require a knowledge of Greek; texts are given in German.

W. K. C. Guthrie (1967)

## PYTHAGORAS AND PYTHAGOREANISM [ADDENDUM 1]

Scholarship on Pythagoras and early Pythagoreanism has undergone a revolutionary change in recent decades. On the one hand, we know much less about Pythagoras and the early school than seemed to be the case a generation ago. On the other hand, it is no longer true that, as W. K. C. Guthrie writes in the original article above, "there is no Pythagorean literature before Plato." Both changes are due to the work of Walter Burkert (1962/1972).

### THE NEW SKEPTICISM ABOUT EARLY PYTHAGOREAN PHILOSOPHY

There had always been skeptics who doubted the traditional view of scientific work by Pythagoras and his early followers. Burkert showed decisively how far this tradition derived from a completely unhistorical view of

Pythagoras created in Plato's Academy and popularized by Plato's immediate successors. The striking similarities between Plato's work and the traditional account of Pythagorean philosophy (as given in Guthrie's article) are largely due to this post-Platonic tradition, which in later versions regularly credited Pythagoras with the invention of Platonic philosophy. At the same time, Burkert defended the authenticity of most of the fragments attributed to Philolaus (in the middle or late fifth century BCE), which are now generally recognized as the earliest Pythagorean texts.

Except for the fragments of Philolaus and a single reference in Plato (reporting that the Pythagoreans regarded music and astronomy as "sister sciences" [*Republic* 530d]), there is no account of Pythagorean philosophy before Aristotle. Furthermore, there is no evidence of any Pythagorean writing before Philolaus. So Aristotle, writing a century and a half after Pythagoras's death, was entirely dependent on an oral tradition for information about the teachings of the early pre-Philolaus school. Since Pythagoras became a figure of legend almost in his own lifetime, it is extremely difficult to know how much of Aristotle's account can be traced back to the founder or to his early followers. We do have some early references to Pythagoras, notably by Heraclitus, but these references are hostile and open to diverse interpretations.

#### PYTHAGOREAN PHILOSOPHY DOWN TO THE TIME OF PLATO

There is no fully reliable account of early Pythagorean doctrine. All we know for certain is that he advocated some version of reincarnation, since Xenophanes, a contemporary, makes fun of Pythagoras for the belief that a human psychē could be reborn in an animal body. Good sources report that Pythagoras founded a cult society or sect, with special dietary restrictions (ultimately including vegetarianism), whose members played an important political role in the cities of southern Italy for several generations. Burkert regarded Pythagoras as essentially a religious teacher, a charismatic guru who founded a ritual community without scientific or philosophic content. Philolaus in the late fifth century then appears as the first Pythagorean philosopher. This interpretation has been followed by many scholars, including the influential study *Philolaus of Croton* by Carl Huffman (1993). On the other hand, the surviving fragments of Philolaus do not show him to be a profoundly original thinker, and some scholars (including the present author) would regard Philo-

laus as formulating, and perhaps updating, an older Pythagorean world view.

The references by Heraclitus to Pythagoras as a polymath who "pursued inquiry further than anyone else" suggest that the more archaic features of this world view go back to the founder himself. As a native of Samos, Pythagoras may well have absorbed the new naturalistic cosmology being worked out in neighboring Miletus during his lifetime. The musical elements in the Pythagorean scheme would then be the personal contribution of Pythagoras himself. It is probably from this school that Plato has taken the notion of the music of the spheres, the cosmic harmony produced by the movement of the heavenly bodies. The numerical proportions corresponding to the musical consonances (2:1, 3:2, 4:3) are embedded in the Sacred Tetractys, the number 10 as the sum of  $1+2+3+4$ , said to be the source of natural order and to provide the oath "by which the Pythagoreans swear." (Diels-Kranz, *Fragmente der Vorsokratiker* 58.B 15) Teachings of this kind may well go back to the first generation in the sixth century. The reports concerning Hippasus of Metapontum (probably dated in the early fifth century) indicate that an interest in the harmonic mean is older than Philolaus. How much work in mathematics or astronomy can be attributed to the early Pythagoreans is another question. But at least one contemporary scholar (Leonid Zhmud 1997) has argued in favor of the traditional view of Pythagoras as founder of a scientific school.

Aristotle reports that the Pythagoreans were pioneers in mathematics and interpreted the whole universe in terms of numbers. However, except for vague reports concerning Hippasus, we know nothing of Pythagorean mathematics in the early period. (The first author to refer to the "Pythagorean theorem" is Plutarch, in the Roman period.) Documentation for Pythagorean teaching begins with Philolaus. He claims that the *kosmos*, or world order, is composed of two opposing principles that are harmoniously fitted together, the principle of Limit or Limiting (*perainonta*) and the Unlimited (*apeiron*). The world is knowable because it is structured by number, and the numbers of special interest are the musical ratios: 2:1, 3:2, 4:3. Aristotle's report of a Pythagorean cosmology in which the earth is a planet like the sun, circulating around a central Fire, seems to be derived from Philolaus. The two Philolaic principles, the Limit and the Unlimited, show up in a famous passage in Plato's *Philebus*, Plato's dialogue, where they are said to have been tossed down from heaven, together with fire, by a certain Prometheus (*Philebus* 16c). Many readers have found it natural to

identify this Prometheus with Pythagoras. This passage may be one of the sources for the ancient story that Plato borrowed his philosophy from Pythagoras.

There is a kernel of truth in the myth that Plato's philosophy is Pythagorean. A genuinely Pythagorean view of the soul as transcending its existence in a human body serves as point of departure for Plato's *Phaedo*. A similar view of the soul is presupposed by the doctrine of recollection in the *Meno* and by the myths of judgment and preexistence in the *Phaedo*, *Republic*, and *Phaedrus*. Furthermore, the other typically Pythagorean view, the mathematical interpretation of nature and the conception of the cosmos in terms of musical harmony, finds its fullest expression in Plato's *Timaeus*. These two dialogues, the *Phaedo* and the *Timaeus*, form the channel through which Pythagorean ideas have passed into the mainstream of ancient and modern thought.

It is important, however, not to exaggerate Plato's debt to the Pythagorean tradition. In most cases, for instance in the judgment myths, Plato has probably transformed Pythagorean material beyond recognition. The pre-Platonic version would appear crude and primitive by comparison. This is perhaps clearest in the case of recollection. Plato's doctrine takes for granted the Pythagorean view of a cycle of reincarnation for the human psyche. But in the original Pythagorean version, recollection would refer only to Pythagoras' alleged ability to recall his previous incarnations, or perhaps to the soul's need to remember certain ritual instructions for correct behavior in the next world—the need to preserve a memory after death that avoids or survives the drink from the River Lethe (forgetfulness). The notion of recollection as a theory in epistemology, as a priori knowledge preceding sensory experience, is entirely Plato's invention. There is no Pythagorean epistemology, and nothing corresponding to the doctrine of Forms. Plato's theory of recollection represents an allegorical reinterpretation of Pythagorean themes that are originally magical or mythical.

In the case of mathematics, however, Pythagorean influence on Plato may be more substantial. The leading Pythagorean of Plato's day, Archytas of Tarentum, was both Plato's friend and also a great mathematician. Of course, not all the mathematics in Plato's dialogues needs to be derived from Pythagorean sources. (Theaetetus, for example, was not a Pythagorean, nor was his teacher Theodorus.) But Plato does cite the Pythagoreans for their view of the relation between music and astronomy (*Republic* 530d, in what is apparently a quotation from Archytas). Unfortunately, we are very poorly informed on

the details of Archytas's work in astronomy and applied mathematics. Hence we cannot tell to what extent his thought inspired Plato's use of geometrical figures and numerical proportions in the cosmology of the *Timaeus*. The role of musical harmonies in the creation of the world soul is at least Pythagorean in spirit. It was the cosmology of the *Timaeus* that became the model for a Pythagorean world view in later centuries, down to the time of Johannes Kepler.

#### PYTHAGOREANISM AFTER PLATO

Aristotle reports that among Plato's "unwritten doctrines" was a theory of two fundamental principles: the One and the Indeterminate Dyad. These principles are regularly attributed to Pythagoras in the post-Aristotelian doxography, together with a mathematical cosmology based on the *Timaeus*. This grandiose and completely unhistorical picture of Pythagorean philosophy, which was accepted throughout antiquity, seems to have been created by Plato's disciples in the Academy and, in particular, by Speusippus, Plato's successor as head of the school. Speusippus composed a book *On Pythagorean Numbers*, which is largely devoted to the cosmological implications of the number 10. This number generates all things by containing as parts the numbers 1, 2, 3, and 4, corresponding, respectively, to point, line, plane, and solid. Thus, Pythagorean numerology, the doctrine of the symbolical and allegorical significance of the numbers from 1 to 10, seems to have begun with Speusippus.

The personal prestige of Pythagoras remained high in the Hellenistic age, although the Pythagorean school seems to have died out by the end of the fourth century BCE. What we find instead are pseudepigraphic works, treatises claiming as their author was either Pythagoras himself or, more frequently, one of his followers. The most popular author for these forged books is Archytas, but there is also a work ascribed to Timaeus of Locri, the fictitious speaker in Plato's dialogue of that name. This treatise has been preserved intact because it was falsely believed to be the Pythagorean original from which Plato derived his cosmology. In general, these pseudepigraphic works contain doctrines borrowed from Plato and Aristotle, with little or no material that is authentically Pythagorean.

A more genuine Pythagorean revival begins in the first century BCE in Rome with a famous Roman magus named Nigidius Figulus, to whom Cicero dedicated his translation of the *Timaeus*. Combining Oriental lore with Greek wisdom and the gift of second sight, Nigidius seems to have created a kind of Pythagorean society

within the Roman aristocracy. The distinguished scholar Marcus Terentius Varro was attracted by this archaic cult, and Cicero himself showed great respect for the Pythagorean tradition. Pythagoras had always been popular in Rome. At the beginning of the third century BCE, in response to an oracle during the Samnite war, a statue of Pythagoras was erected in the Roman forum. The popularity of Pythagoras in Rome was due in part to the fact that the original Pythagorean community had been located in the Greek cities of southern Italy, and hence Pythagoreanism was known since Aristotle as the *Italian philosophy*. Since most philosophers were from Greece proper, the Romans were pleased to have their own sages from Croton, Metapontum, and Tarentum.

At the same time that Nigidius was reviving Pythagoreanism in Rome in the first century BCE, Eudorus of Alexandria was inaugurating a new line of Platonic philosophers who have come to be called the Neopythagoreans. The name reflects the fact that these philosophers share the view that Plato's philosophy was derived from Pythagoras. Thus, Eudorus reports that "Socrates and Plato agree with Pythagoras that the goal of life (the *telos*) is becoming like god. But Plato articulated it more clearly by adding 'as far as possible.'" (Plato, *Theaetetus* 176b.) In reaction against the tradition of Skepticism that prevailed in the Hellenistic Academy, these philosophers emphasize the metaphysical and theological elements in Plato's philosophy. In this respect, and in the central importance attributed to numbers, the Neopythagoreans return to a position like that of Speusippus and Xenocrates in the Old Academy.

According to Eudorus, the Pythagoreans regard the One as the first principle of all things and the supreme god, but immediately below it are the two opposed principles, the One and the Indefinite Dyad. This notion of a hierarchical system of transcendental principles was developed in a new version of Pythagorean philosophy by Moderatus of Gades in the first century CE. Here there are three levels of nonsensible reality represented by three Ones. (The three Ones are related to the first three hypotheses in Plato's *Parmenides*.) Other Platonists in this tradition refer to a doctrine of three gods as distinctively Pythagorean. The best-preserved view is that of Numenius of Apamea in the second century CE. The first god is pure *nous*, an intellect focussed only on itself, like Aristotle's Prime Mover; the second god is *nous* as the demiurge, responsible for ordering the material universe; the third god is either the visible cosmos or its animating principle, the world soul. There is a significant parallel between this tripartite scheme and the three hypostases of

Plotinus (the One, the Intellect, and the Soul), and it is not surprising to learn that Plotinus was accused of borrowing his philosophy from Numenius.

Pythagorean influence continued into late antiquity and the middle ages, both as numerology and as integrated into Neoplatonism, above all in the work of Iamblichus (c. 300 CE), who composed a book *On the Pythagorean Way of Life* as an introduction to his major work *On the Pythagorean School*. Pythagorean harmonics (through the influence of Boethius) continued to play a role in music down to the modern age. Finally, in the Renaissance, Pythagorean ideas were revived with the new access to Plato and the Neoplatonists. This leads, on the one hand, to a flowering of occult numerology and theosophy—for example, in the cosmology of Robert Fludd—and on the other hand, to scientific applications of Pythagorean thought in the work of Copernicus and Kepler. It was Kepler who made the last great scientific contribution to the Pythagorean tradition. Taking as his model the *Timaeus* and Ptolemy's *Harmonica*, Kepler published his laws of planetary motion in a work titled *Harmonice Mundi* (The harmonics of the universe), in which he undertook to show how the movements of the planets were designed to illustrate the Pythagorean music of the spheres. Kepler's work brings the story of scientific Pythagoreanism to a happy conclusion. In one sense, the spirit of Pythagoreanism is still alive today in the mathematical interpretation of nature; string theory may be the latest version of the harmony of the spheres.

**See also** Archytas of Tarentum; Aristotle; Boethius, Anicius Manlius Severinus; Cicero, Marcus Tullius; Copernicus, Nicolas; Epistemology; Fludd, Robert; Hellenistic Thought; Heraclitus of Ephesus; Iamblichus; Kepler, Johannes; Numenius of Apamea; Philolaos of Croton; Philosophy of Science, History of; Plato; Plotinus; Plutarch of Chaeronea; Renaissance; Xenophanes of Colophon.

### Bibliography

- Burkert, W. *Weisheit und Wissenschaft: Studien zu Pythagoras, Philolaos, und Platon*. Nürnberg: H. Carl, 1962. Revised translation by Edwin L. Minar, *Lore and Science in Ancient Pythagoreanism*. Cambridge, MA: Harvard University Press, 1972.
- Dillon, J. *The Middle Platonists*. Ithaca, NY: Cornell University Press, 1977.
- Huffman, Carl A. *Archytas of Tarentum: Pythagorean, Philosopher, and Mathematician King*. Cambridge, U.K.: Cambridge University Press, 2005.

- Huffman, Carl A. *Philolaos of Croton: Pythagorean and Presocratic*. Cambridge, U.K.: Cambridge University Press, 1993.
- Huffman, Carl A. "The Pythagorean Tradition." In *Cambridge Companion to Early Greek Philosophy*, edited by A. A. Long. Cambridge, U.K.: Cambridge University Press, 1999, pp. 66–87.
- Hussey, E. "Pythagoreans and Eleatics." In *Routledge History of Philosophy*, Vol. 1: *From the Beginnings to Plato*, edited by C. C. W. Taylor. London: Routledge, 1997, pp. 128–174.
- Kahn, C. H. "Pythagorean Philosophy Before Plato." In *The Presocratics. A Collection of Critical Essays*, edited by Alexander P. D. Mourelatos. Princeton, NJ: Princeton University Press, 1993, pp. 161–185.
- Kahn, C. H. *Pythagoras and the Pythagoreans*. Indianapolis, IN: Hackett, 2001.
- Riedweg, C. *Pythagoras. Leben, Lehre, Nachwirkung*. Munich: C.H. Beck, 2002.
- Zhmud, L. *Wissenschaft, Philosophie und Religion im frühen Pythagoreismus*. Berlin: Akademie Verlag, 1997.

Charles H. Kahn (2005)

## PYTHAGORAS AND PYTHAGOREANISM [ADDENDUM 2]

Ideas of Pythagoras and his school (including Philolaos) became known to the Islamic and to a lesser degree to the Jewish world since the end of the ninth century. Doxographical information about them can be found in Arabic translations of Aristotle, Plato, and above all two doxographical sources: the *Placita philosophorum*, which is attributed to Plutarch and is assumed to be compiled by Aetius Arabus (Daiber 1980), and a doxography that is attributed to Ammonius and is available only in an Arabic version (Rudolph 1989), much like the Arabic translation of the *Placita* apparently from the second half of the ninth century.

The impact of these sources, especially of Aetius, on Islamic thought (Rosenthal 1965, Daiber 1980), p. 337f.), on the Islamic philosopher al-Kindī, who died in 866 (Baffioni 1985), on the anonymous encyclopaedia of the Sincere Brethren from the tenth century (Netton 1991), and on the Jewish philosopher Sa'adia ben Joseph in the first half of the tenth century (Efros), was concentrated on the Pythagorean doctrine of numbers, especially of the number four as source of the cosmos and its harmony, and also applied to music. Shahrastani's exposition of the Pythagorean doctrine (Baffioni 1983), pp. 96ff.), which mainly combined the reports of Aetius and in Ara-

bic doxography attributed to Ammonius (Rudolph 1989), shaped the picture of Pythagoras among Islamic thinkers.

Moreover, Neo-Pythagorean texts on ethics contributed to the propagation of Pythagorean thought in Islamic and Jewish circles. Here, an important role was played by the Pythagorean *Carmina aurea* on ethical principles of life such as piety, modesty, justice, and self-examination as ways of the soul's assimilation to God. This text was known to the Arabs in an anonymous Arabic translation from the second half of the ninth century, which was integrated in Ḥunayn ibn Ishāq's *Nawādir al-falāsifa* (Anecdotes of the Philosophers), a collection of wise sayings that was often used by Muslim authors (Baffioni 1994, Miskawayh 1964), and that in the adaptation of Muhammad ibn 'Alī al-Anṣārī was translated into Hebrew (Daiber 1995).

Originally, the *Carmina aurea* were translated into Arabic with the commentary by Iamblichus (250–330 CE), a pupil of the neoplatonic philosopher Porphyry. This commentary, which is lost in its Greek original and preserved in Arabic (Daiber 1995), differs from that attributed to Proclus, which in a similar manner offers Neo-Pythagorean traditions in neoplatonic shape and which is preserved in a redaction by Abū l-Faraj ibn al-Tayyib from the eleventh century (Linley 1984; cf. Daiber, *Islam* 65 1988, 134–137). Iamblichus's commentary continues the discussion of his *De vita pythagorica* and *Protrepticus* and amalgamates Pythagorean, Platonic-neoplatonic, Aristotelian, and Stoic ethics. It found an echo in al-Kindī, who in his "Summary on the Soul According to Aristotle, Plato, and Other Philosophers" describes the ascent and return of the soul to its divine origin through purification and increasing knowledge of God—a doctrine that is developed a century later in the encyclopaedia of the Sincere Brethren (Baffioni 1992) and is alluded to in Ibn Sīnā's (Avicenna) (d. 1037) alleged Pythagoreanism (Chaix-Ruy 1959).

Iamblichus's neoplatonic tradition of the *vita pythagorica* is reflected in a treatise attributed to Plato, "The Exhortation concerning the Education of Young Men," which is preserved only in Arabic (Rosenthal 1941, pp. 383ff.). It can be traced back to his teacher, Porphyry, who in his *History of Philosophy* (lost in Greek and preserved in some Arabic fragments) had included the biography of Pythagoras (Rosenthal 1990). It seems plausible that the same neoplatonic tradition of the *vita pythagorica* also affected the alleged letter by Pythagoras to Hiero, the tyrant of Sicily, which is available in a clumsy ninth-century translation (Rosenthal 1975). Finally, Neo-Pythagorean ethics is mirrored in the numerous sayings



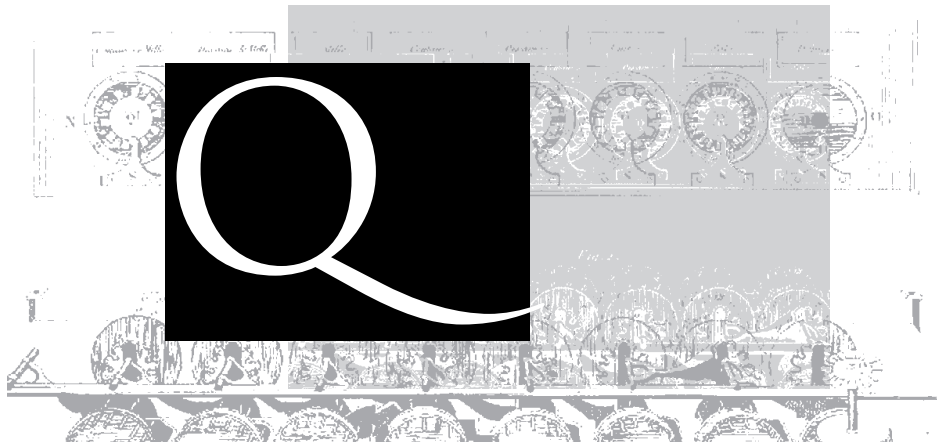
attributed to Pythagoras and transmitted in Syriac and Arabic gnomologia (Gildemeister 1870, Levi della Vida 1910, Gutas 1975).

### **Bibliography**

- Baffioni, Carmela. “‘Detti aurei’ di Pitagora in trasmissione araba.” *I moderni ausili all’Ecdotica. Atti del Convegno Internazionale di Studi* (Fisciano–Vietri sul Mare–Napoli, 27–31 ottobre 1990). A cura di Vincenzo Placella e Sebastiano Martelli. Napoli. 1994 (= Pubblicazioni dell’Università degli studi di Salerno. Sezione atti, convegni, miscellanee. 39): 107–131
- Baffioni, Carmela. “Platone, Aristotele e il pitagorismo kindiano.” *Annali dell’Istituto Universitario Orientale di Napoli. Pubblicazioni* (Napoli) 45 (1985): 135–144.
- Baffioni, Carmela. “Traces of ‘Secret Sects’ in the Rasa’ il of the Ikhwan al-Safa’.” In *Shi‘a Islam, Sects, and Sufism: Historical dimension, religious practice and methodological considerations*. Edited by F. DeJong. Utrecht: M. Th. Houtsma Stichting, 1992, 10–25.
- Baffioni, Carmela. “Una ‘storia della filosofia greca’ nell’islam del XII secolo: III. Pitagora.” *Elenchos* (Napoli) 4 (1983): 93–132, 261–345.
- Chaix-Ruy, J. “Du pythagorisme d’Avicenne au soufisme d’Al-Ghazali.” *Revue de la mediterranneé* (Alger) 19 (1959): 289–327.
- Daiber, Hans. *Aetius Arabus: Die Vorsokratiker in arabischer Überlieferung*. Wiesbaden, Germany: Steiner, 1980.
- Daiber, Hans. *Neuplatonische Pythagorica in arabischem Gewande: Der Kommentar des Iamblichus zu den Carmina aurea*. New York: North-Holland, 1995.
- Efros, Israel Isaac. “Saadyah’s Second Theory of Creation in Its Relation to Pythagoreanism and Platonism.” *Louis Ginzberg Jubilee Volume English Section* (1945–1946): 133–142.
- Gildemeister, J. “Pythagorassprüche in syrischer Überlieferung.” *Hermes* 4 (1870): 81–98.
- Gutas, Dimitri. *Greek Wisdom Literature in Arabic Translation: A Study of the Graeco-Arabic Gnomologia*. New Haven, CT: American Oriental Society, 1975.
- Levi della Vida, Giorgio. “Sentenze pitagoriche in versione siriana.” *Rivista degli studi orientali* 3 (1910): 595–610.
- Linley, Neil, ed. and trans. *Ibn at-Tayyib: Proclus’ Commentary on the Pythagorean Golden Verses*. Buffalo, NY: Department of Classics, State University of New York, 1984.
- Miskawayh. *An Unpublished Treatise of Miskawayh on Justice or Risala fi mahiyyat al-‘Adl ‘i Miskawayh*, edited and translated by M. S. Khan. Leiden, Netherlands: E. J. Brill, 1964.
- Netton, Ian Richard: *Muslim Neoplatonists: An Introduction to the Thought of the Brethren of Purity (Ikhwan al-Safā’)*. Edinburgh: Edinburgh University Press, 1991.
- Rosenthal, Franz. “Arabische Nachrichten über Zenon den Eleaten.” In *Greek Philosophy in the Arab World: A Collection of Essays*. Aldershot, U.K.: Grower Publishing, 1990. Rosenthal’s essay originally published in the journal *Orientalia* N.S. 6 (1937): 27–67.
- Rosenthal, Franz. “Fithaghuras.” In *Encyclopaedia of Islam* 12, 929–930. Leiden, Netherlands: E. J. Brill, 1965.
- Rosenthal, Franz. “From Arabic Books and Manuscripts, XIII–XIV.” *Journal of the American Oriental Society* 95 (2) (1975): 209–213.
- Rosenthal, Franz. “Some Pythagorean Documents Transmitted in Arabic.” *Orientalia* 10 (1941): 104–115, 383–395.
- Rudolph, Ulrich. *Die Doxographie des Pseudo-Ammonios: Ein Beitrag zur neuplatonischen Überlieferung im Islam*. Stuttgart, Germany: Steiner, 1989.

**Hans Daiber (2005)**





## QUALIA

The word *quale* (or *qualia*) derives from the Latin for “quality.” As used by C. I. Lewis (1929) and those following him, it refers to the qualities of phenomenal individuals, such as color patches, tastes, and sounds. In this sense the term means what George Berkeley meant by “sensible qualities,” or what later philosophers meant by *sensa* or *sense data*. Since the demise of sense data theories, the term *qualia* has come to refer to the qualitative, or phenomenal, character of conscious, sensory states, so that it is mental states, not phenomenal individuals, that are the subjects of predication. Another expression for this aspect of mental life is the “raw feel” of experience, or “what it’s like” to have certain sensory experiences. Qualia are part of the phenomenon of the subjectivity of consciousness, and pose one of the most difficult problems for a materialist solution to the mind-body problem.

### IDENTITY THEORY

J. J. C. Smart posed the challenge this way in a 1959 article: Consider a sensation like a yellowy-orange after-image. According to the materialist theory known as the “central state identity theory” (or just “identity theory”), the sensation is a brain state. Smart’s worry, which he

attributed to Max Black, was that even if one accepted that the sensation was itself a brain state, it still seemed as if one had to attribute an “irreducibly psychic” property to the brain state. That is, there is a distinctive qualitative character experienced when having a yellowy-orange after-image, and that property—that yellowy-orange character—does not seem at all like a physical property. So even if all mental states are brain states, we might still be driven to the view that some mental properties—qualia, in particular—are not physical. This would constitute a form of dualism known as “property dualism,” a position inconsistent with materialism.

The Max Black objection presented by Smart in 1959 is related closely to the “conceivability argument,” a dualist argument going back to René Descartes, and revived in 1980 by Saul Kripke and in 1996 by David Chalmers. Roughly, the idea is this. When one considers simultaneously what it is like to see a yellowy-orange after-image and any description of the firing pattern of an assembly of neurons, it seems perfectly coherent to imagine having the one without the other. That neurons should fire in this or that pattern and that it should be like nothing at all for the subject whose neurons they are seems clearly possible. Yet, if qualia are identical to neural properties, such a situation is not possible. Hence, qualia must be not

neural properties, but nonphysical properties possessed by neural states.

Another closely related dualist argument is Frank Jackson's 1982 "knowledge argument." We are asked to imagine a scientist who knows everything about the physiology of color vision, but who has never seen anything in color. Upon first seeing a red rose, it seems clear that she would learn something new—what it is like to see red. Yet, if qualia are just physical properties of the nervous system, she should have already known what it is like to see red. Hence, Jackson concludes, qualia are not physical properties.

Many materialist philosophers object that these dualist arguments rely on an assimilation of concepts and properties. Concepts are elements of thought, ways of thinking of objects and properties, comparable to words in a language. Just as there can be many distinct words referring to the same object or property, so too there can be distinct concepts that apply to the same property. All the above arguments demonstrate, according to these philosophers, is that we have different ways of conceiving of qualia, and that it isn't obvious that they pick out the same properties. But just as the fact that we had to learn that water is identical to H<sub>2</sub>O does not impugn the claim that they are identical, so too the fact that we have to learn that a certain quale is a certain neural property does not refute the claim that they are indeed the very same property.

### EXPLANATORY GAP

Proponents of property dualism respond that there are important differences between the water–H<sub>2</sub>O case and the case of qualia that undermine the analogy pushed by materialists (Chalmers 1996). However, even if the conceivability and knowledge arguments do not demonstrate that qualia are, as Smart put it, "irreducibly psychical," they do point toward another problem, one that goes under the name of the "explanatory gap" (Levine 1983, 2001). The problem is this. If qualitative sensory experiences are really nothing over and above the interplay of neural firing in the relevant part of the brain, then one would expect that the qualitative character of particular types of sensations could be explained and predicted by reference to their neurophysiological embodiments. Yet, when we consider what it is like to see a red rose or a yellowy-orange after-image, it seems completely arbitrary that it should be the result of this type of neural firing as opposed to some other. In fact, it seems totally arbitrary that it should be like anything at all, merely from a knowledge of the neural properties. In this sense there seems to

be an explanatory gap between the underlying level of neurophysiological phenomena and the level of qualitative experience. Thomas Nagel (1974) makes a similar argument about the limits of materialist understanding by noting that as much as we learn about the echolocation sense of bats, we can never learn thereby what it is like to be a bat and to sense the world in this way.

Faced with these strong intuitions that there is something suspect about the connection between physical properties of the nervous system and qualia, materialists have adopted two different strategies. The first is to attempt to straightforwardly dispel these anti-materialist intuitions by coming up with materialist theories of qualia that are intuitively acceptable. The second is to grant the apparent mystery involved in the connection between qualia and neurophysiological properties, but to argue that there are reasons why this connection should appear so mysterious that do not in the end contradict the basic tenets of materialism.

In line with the first strategy, Smart (1959) himself addressed the problem by proposing what he called a "topic-neutral" analysis of qualitative character. His claim was that our notion of qualitative character is neutral with respect to the kind of material in which it is embodied. Rather, to have a sensation of a yellowy-orange after-image, for instance, is to occupy a state that is similar to the state one is in when actually seeing an orange. This idea was then later developed by functionalists such as H. Putnam (1991), who identified mental states of all kinds with causal roles. That is, a particular mental state is defined as a state that is caused in certain characteristic ways (by physical stimuli and other mental states) and has certain characteristic effects (particular forms of behavior as well as other mental states). On this view the connection between a particular qualitative sensory state and a brain state is truly contingent, since it is allowed that any other physical state that filled the same causal role would count as an instance of this qualitative state.

Adopting functionalism for qualia might seem to provide the materialist with a response both to the conceivability argument and to the problem of the explanatory gap. It is conceivable that a creature might experience a certain sensory quality without being in a particular physical state since there are many different physical states that can support the relevant causal role. Also, one might occupy a certain physical state without having the sensory experience because that state is not connected in the right way to other states, and therefore is not playing the appropriate causal role. As for the explanatory gap, the idea is that appeal to the intrinsic

physical properties of a brain state don't explain its mental, or qualitative character, because qualitative character is a function of the relations that physical state maintains with other internal physical states, as well as stimuli and behavior. The proper locus of explanation for the qualitative character of experience is the overall pattern of interactions among the subject's internal states; it is a matter of the structure, not the "stuff" in which the structure is embodied.

However, it turns out that almost the very same problems that attended the identity theory return to haunt functionalism as well. Take the conceivability argument. In the form of the "inverted qualia" and "absent qualia" hypotheses the conceivability argument can be mounted against functionalism as well. The inverted qualia hypothesis is the conjecture that there could be two functionally identical creatures—that is, both creatures, though made of different material, possess a set of internal states that maintain the very same pattern of interactions with each other and the relevant inputs and outputs to the system—that experience very different qualia when occupying the very same functional state.

The standard illustration of this possibility is known as the "inverted spectrum hypothesis." Oversimplifying greatly for now, consider the fact that the color wheel can be inverted in such a way as to maintain all of the similarity relations. That is, if one creature sees blue and green where another sees yellow and red, and vice versa, then all of their judgments about the relative similarities of objects with respect to color would converge. Imagine that this inversion occurred at birth, so they learned to use color terms the same way. Jack and Jill might both call a ripe tomato red, though Jack's experience is qualitatively like what Jill would experience were she looking at a ripe cucumber.

If such an inversion of qualia with respect to functional roles is possible, then the qualitative character of a sensory experience cannot be identified with its functional role. To make matters worse, it seems perfectly coherent to imagine a creature that satisfies the relevant functional description, and yet for whom there is no conscious experience occurring at all. (Often such creatures are known as "zombies" in the literature.) Ned Block (1980) describes a very compelling example. Imagine, he says, the entire nation of China connected by phone lines in such a way that, collectively, they satisfy the same functional description as a human brain. Would we want to say that the entire nation of China, as a single subject, is seeing red, or feeling pain? Certainly it seems at least possible that no genuine experience is going on at all. Hence

having a qualitative sensation cannot be merely a matter of possessing internal states that play a certain functional, or causal role.

## THE DILEMMA

The objections to both the identity theory and functionalism reveal a deep dilemma for materialists about qualia. The qualitative character of a sensation—the way color looks, the way pain feels—strongly seems to be an intrinsic property of the sensation, a matter of how things are with one at that moment, not a matter of how one is disposed to act or what effects are likely. In this sense the identity theory seems quite appropriate, since it identifies the qualitative character of an experience with a physical property of the brain state one occupies at that moment. The problem is that there seems to be no intelligible connection between the physical properties of brain states and the qualitative properties of sensory experiences.

If, however, we pin qualitative character on the pattern of relations that a sensory state maintains with other states, as well as stimuli and behavior, then we can see how appeal to the physical properties of brain states could play an important explanatory role. We can explain how it is that the neural state one occupies when, say, experiencing a yellowy-orange after-image, interacts with other neural states and stimuli and behavior so as to realize the relevant pattern by appeal to the causal mechanisms of the brain and nervous system. The only problem here is that, as demonstrated by the inverted and absent qualia hypotheses, qualitative character is not convincingly characterizable as a matter of the pattern of interactions among internal states. Thus the materialist is faced with this dilemma: Qualitative character is explicable in physical terms only if it can be characterized as a pattern of causal relations among mental states, but only a theory of qualitative character that treats it as an intrinsic property of mental states will be intuitively acceptable.

Functionalism is a structural theory of qualitative character—a particular quale is identified with a particular niche in the overall system of causal interactions among stimuli, internal states, and behavior. Another structural theory worthy of mention is what we might call the "quality space" theory, proposed by Austen Clark in 1993. On this view we start with the idea that different sensory modes—vision, hearing, etc.—define quality spaces. A quality space is a multidimensional space whose axes are determined by the number of independent parameters along which sensory experiences in a given mode can vary. To take again an admittedly oversimplified example, consider color vision. Colors vary along

three dimensions: hue, brightness, and saturation. A particular color (where this means a determinate shade) can then be identified with a vector representing its values in each of the three dimensions. Assuming colors only vary in these three ways (which is part of the oversimplification), then a person's similarity judgments about colors can be predicted and explained by the distances among the relevant color vectors. A complete map of a sensory system can be drawn once all of the independent parameters of variation have been determined. Qualia, then, are points in quality spaces.

The quality space view differs from functionalism in that the structure by reference to which a quale is defined is not a pattern of causal interactions, but rather a quality space. However, it shares with functionalism the idea that it is structural relations rather than intrinsic features of the experience that determine qualitative character. Also, like functionalism, on the quality space view the appeal to the physical features of neural states comes in to explain how the relevant structure is embodied. This allows for the possibility of many alternative physical embodiments for the same quality space, so long as the overall structure of relations among the elements is preserved. Unfortunately, also like functionalism, the quality space view is subject as well to the problems of the inverted and absent qualia hypotheses. So with respect to the general dilemma facing the materialist it does not improve on functionalism.

## REPRESENTATIONALISM

A view that seems to promise a way to overcome the materialist's dilemma is "representationalism," discussed in the work of Fred Dretske (1995), G. Harman (1990), William Lycan (1987), and Michael Tye (1995). One way to motivate the theory is to start from an untenable but nevertheless quite tempting theory of qualia and then see representationalism as a way to capture the spirit of the original view while removing its fatal weakness. The tempting but untenable view is this. Qualia, rather than features of mental states, are properties of external objects. They are the colors, sounds, and textures out there in the world that our senses detect. This view is tempting for two reasons. First, it removes qualia as obstacles to a materialist solution to the mind-body problem, since qualia are no longer features of mental states. Second, it is intuitively plausible. Advocates of the view often defend it by citing the so-called transparency of sensory experience. If asked to describe what it is like to have various sensory experiences, one finds oneself describing the properties of external objects. One says

things like "it looks like a lemon," "it tastes like chicken," or "it feels smooth."

The reason the view is untenable is that it cannot handle cases of hallucination or illusion. Suppose one "sees" a pink elephant where there is nothing remotely pink or elephant-like. Clearly one is having a sensory experience with a "pinkish" qualitative character, yet there is no object out there in the world that is pink. Hence the quale cannot be the pink of the elephant, it has to be a feature of one's experience, a property of one's internal mental state. Representationalism comes into play at this point. According to this view, sensory states are mental states that represent the way the world is around us. They differ from belief states in being nonconceptual, more picturelike, but they share with beliefs and thoughts the feature of representing the world. Qualia, then, are the representational contents of sensory experiences. That is, to have a "pinkish" qualitative character is for one's visual state to have the content that something out there in the world is pink. Notice that representationalism shares with the original view the core idea that pinkness is primarily a feature of external objects, but it nevertheless accommodates hallucination and illusion. Just as one can think that there are pink elephants even though there aren't any, so too one can have visual experiences that represent pink elephants even though there aren't any.

While representationalism has many virtues, there are two primary problems. First, the view is less plausible when applied to bodily sensations like pains and itches than when applied to colors and sounds. What does the qualitative character of an itch or a headache represent? Advocates of representationalism maintain that these sensations represent conditions of the body. Whether this view can be sustained is a matter of controversy. The main problem, however, is that representationalism does not overcome the basic challenge facing other materialist theories. Just as functionalism and quality space theory have trouble with inversion and zombie scenarios, so too does representationalism. It seems easy to imagine a creature who normally sees objectively red objects the way others see objectively green objects, and it also seems possible for there to be creatures, or devices, that meet the relevant specifications for representing the qualities of external objects in a "sensory" format but for whom there is nothing it is like to occupy these representational states. Properly programmed computers certainly seem like possible examples. Hence the principal challenge to materialist theories of qualia remains.

## PHENOMENAL CONCEPTS

Some materialist philosophers dismiss the inverted and absent qualia hypotheses, along with the conceivability argument, by insisting that the intuitions that underlie these challenges are just that—intuitions—and should not be accorded much significance. Daniel Dennett (1988) and Georges Rey (1997) go so far as to embrace eliminativism, the view that qualia do not really exist. We think we have these features of experience, but in fact they represent a kind of cognitive illusion. However, other materialists, such as Brian Loar (1997), William Lycan (1987), Colin McGinn, David Papineau (2002), Scott Sturgeon (2000), and Michael Tye (1995), insist on the reality of qualia and grant the import of the intuitive resistance to materialism. Their strategy is to attempt to provide a satisfactory materialist theory of the intuitive resistance itself. For many the main tool in this endeavor is the notion of a “phenomenal concept.”

Phenomenal concepts are the special concepts of qualitative properties that we employ when thinking of our qualitative states from within the first-person point of view. When one considers what it is like to see a red rose, and then says something like, “How can *that* be merely a matter of neurons firing in a certain pattern?” one is employing a phenomenal concept to think about the experience. The proposal then is to explain the stubborn cognitive resistance to materialist theories (of whatever form) by appeal to peculiar features of phenomenal concepts. It is a feature of our cognitive architecture, on this view, that we cannot come to see how the qualitative character of our experience is just a matter of the way our neurons are firing. We are doomed to suffer from an explanatory gap, but that we are so doomed is itself explainable in perfectly respectable materialist terms.

Whether this appeal to phenomenal concepts can do the work of extricating materialism from the challenges posed by qualia is still a matter of controversy. It appears that the cluster of problems comprising consciousness, qualia, and subjectivity are destined to haunt the philosophy of mind for some time to come.

**See also** Consciousness; Knowledge Argument.

### Bibliography

Block, N., and J. A. Fodor. “What Psychological States Are Not.” *The Philosophical Review* 81 (1972): 159–181.

Block, Ned. “Troubles with Functionalism.” In *Readings in Philosophy of Psychology*, edited by Ned Block. vol. 1, 268–305. Cambridge, MA: Harvard University Press, 1980–1981.

Chalmers, David John. *The Conscious Mind: In Search of a Theory of Conscious Experience*. Oxford: Oxford University Press, 1996.

Churchland, Paul M. “Reduction, Qualia, and the Direct Introspection of Brain States.” *Journal of Philosophy* 82 (1985): 8–28.

Clark, Austen. *Sensory Qualities*. Oxford: Oxford University Press, 1993.

Davies, Martin, and Glyn W. Humphreys, eds. *Consciousness: Psychological and Philosophical Essays*. Oxford: Blackwell, 1993.

Dennett, Daniel C. *Consciousness Explained*. Boston: Little, Brown, 1991.

Dennett, Daniel C. “Quining Qualia.” In *Consciousness in Contemporary Science*, edited by A. J. Marcel and E. Bisiach. Oxford: Oxford University Press, 1988.

Dretske, Fred. *Naturalizing the Mind*. Cambridge, MA: MIT Press, 1995.

Hardin, C. L. *Color for Philosophers: Unweaving the Rainbow*. Indianapolis: Hackett, 1988.

Harman, G. “The Intrinsic Quality of Experience.” In *Philosophical Perspectives*, 4, edited by J. Tomberlin, 31–52. Atascadero, CA: Ridgeview, 1990.

Jackson, Frank. “Epiphenomenal Qualia.” *Philosophical Quarterly* 32 (1982): 127–136.

Kripke, Saul A. *Naming and Necessity*. Cambridge, MA: Harvard University Press, 1980.

Levine, Joseph. “Materialism and Qualia: The Explanatory Gap.” *Pacific Philosophical Quarterly* 64 (1983): 354–361.

Levine, Joseph. *Purple Haze: The Puzzle of Consciousness*. New York: Oxford University Press, 2001.

Lewis, Clarence Irving. *Mind and the World-Order: Outline of a Theory of Knowledge*. New York: Scribner’s Sons, 1929.

Loar, Brian. “Phenomenal States.” In *The Nature of Consciousness: Philosophical Debates*, edited by N. Block, O. Flanagan, and G. Güzeldere. Cambridge, MA: MIT Press, 1997.

Lycan, William G. *Consciousness*. Cambridge, MA: MIT Press, 1987.

McGinn, Colin. *The Problem of Consciousness: Essays toward a Resolution*. Oxford: Basil Blackwell, 1991.

Nagel, Thomas. “What Is It Like to Be a Bat?” *The Philosophical Review* 83 (4) (1974): 435–450.

Papineau, David. *Thinking about Consciousness*. Oxford: Oxford University Press, 2002.

Putnam, H. “The Nature of Mental States.” In *The Nature of Mind*, edited by David M. Rosenthal. New York: Oxford University Press, 1991.

Rey, Georges. *Contemporary Philosophy of Mind: A Contentiously Classical Approach*. Cambridge, MA: Blackwell, 1997.

Shoemaker, Sydney. *Identity, Cause, and Mind: Philosophical Essays*. Cambridge, U.K.: Cambridge University Press, 1984.

Smart, J. J. C. “Sensations and Brain Processes.” *The Philosophical Review* 68 (1959): 141–156.

Sturgeon, Scott. *Matters of Mind: Consciousness, Reason, and Nature*. London: Routledge, 2000.

Tye, Michael. *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA: MIT Press, 1995.

**Joseph Levine (1996, 2005)**

## QUALITIES, PRIMARY AND SECONDARY

See *Primary and Secondary Qualities*

## QUALITY AND QUANTITY

See *Categories; Dialectical Materialism; Primary and Secondary Qualities*

## QUANTIFIERS IN FORMAL LOGIC

Familiarity with classical quantification theory is presupposed here. Some proposed amendments are considered, as are several additions.

### ALTERNATIVES TO CLASSICAL QUANTIFICATION THEORY

First-order logic can be reformulated so as to avoid quantifiers and variables. This is only partially done in modal logic, which avoids explicit quantification over possible states of the world in favor of operators  $\Box$  and  $\Diamond$ . However, in principle all quantification is avoidable, if one is willing to admit enough operators and does not worry about their having ordinary-language readings. In practice, however, few have preferred this predicate-functor approach (see Quine 1960, Benthem 1977). Thus, even such dissidents as the intuitionists adopt the classical quantificational language, though the properties they ascribe to the quantifiers are nonclassical. (Thus, while classically  $\forall$  and  $\neg\neg\forall$  and  $\forall\neg\neg$  are equivalent, intuitionistically the first is stronger than the second and the second stronger than the third.)

Classical logic allows terms formed from constants and function symbols, subject to the restriction that each term must denote some element of the domain over which the quantifiers range; but terms are eliminable using Bertrand Russell's theory of descriptions. On the classical Tarskian definition of truth in a model, truth of  $\forall x\phi(x)$  (respectively,  $\exists x\phi(x)$ ) is equivalent to the truth of  $\phi(t)$  for all (respectively, some) terms  $t$  only in special cases, as when each element of the domain is the denotation of some term of the language (which is never so if the domain is uncountable and the language countable). By contrast, the so-called substitutional quantifier  $\prod$  (respectively,  $\Sigma$ ) is defined by the condition that  $\prod x\phi(x)$  (respectively,  $\Sigma x\phi(x)$ ) always counts as true if and only if

(iff)  $\phi(t)$  is true for all (respectively, some) terms  $t$ . There is no technical obstacle to introducing such operators, but whether there is any philosophical advantage to doing so is controversial. In particular, if one has in mind a specific domain,  $\prod$  (respectively,  $\Sigma$ ) will be intuitively equivalent to the ordinary language "for every (respectively, some) element of the domain" only in special cases (see Kripke 1976). Antithetical to substitutional quantification is so-called free logic, which drops the classical restriction that all terms must have denotations and gives up the classical inferences from  $\forall x\phi(x)$  to  $\phi(t)$  and from  $\phi(t)$  to  $\forall x\phi(x)$  (see Bencivenga 1983).

### EXTENSIONS OF CLASSICAL QUANTIFICATION THEORY

In contrast to the various anticlassical logics just mentioned, by far the largest body of work on quantifiers in formal logic concerns certain extraclassical logics, called model-theoretic logics. These accept classical logic and the Tarskian definition of truth in a model, but introduce additional kinds of quantifiers into the language, indicating their intended meaning by adding clauses for them to the Tarskian definition. There are several kinds (see Barwise and Feferman 1985).

**CARDINALITY QUANTIFIERS.** Though there are nineteenth-century and even medieval antecedents, the modern theory of such quantifiers as "most" begins with Andrzej Mostowski (1957). Given a formula  $\phi(x)$  and a model with domain  $A$ , write  $\phi[a]$  to indicate that  $a \in A$  satisfies  $\phi(x)$ ; also write  $\text{card } B$  for the cardinality of a set  $B$ . Then the truth conditions for the most studied Mostowski-style quantifiers are as shown in Table 1.

All these generalized quantifiers count as logical notions according to the definition of Alfred Tarski (1986) (which requires that any sentence involving a purportedly logical operator that is true in a model remains true if the model is replaced by an isomorphic one). Their theory has been worked out in some detail. For example, for first-order logic plus  $Q_0$  the Löwenheim-Skolem theorem holds but the compactness theorem fails, while for  $Q_1$  the opposite is the case.

**PLURAL QUANTIFIERS.** So-called second-order and higher-order quantifiers are nowadays generally read as first-order quantifiers, but with a different domain from that of the first-order quantifiers. Thus, one writes " $\exists X(Xy \ \& \ \dots)$ " but reads it as something like "There is a class  $X$  such that  $y$  is a member of  $X$  and ..." or "There is a concept  $X$  such that  $y$  falls under  $X$  and ..." and simi-



**TABLE 1**

Quantifier	Truth condition
<b>Most</b> $x\phi(x)$	card $\{a: \phi[a]\} > \text{card } \{a: \neg\phi[a]\}$
<b>More</b> $x[\phi(x), \psi(x)]$	card $\{a: \phi[a]\} > \text{card } \{a: \psi[a]\}$
<b>Q<sub>0</sub></b> $x\phi(x)$	card $\{a: \phi[a]\}$ infinite
<b>Q<sub>1</sub></b> $x\phi(x)$	card $\{a: \phi[a]\}$ uncountable
<b>H</b> $x\phi(x)$	card $\{a: \phi[a]\} = \text{card } A$
<b>R</b> $xy\phi(x)$	for some infinite $I \subseteq A$ , $\psi[a, b]$ for all distinct $a, b \in I$

larly for the two-place “ $\exists X(Xyz \ \& \ \dots)$ ” and the third-order “ $\exists X(XY \ \& \ \dots)$ ,” with *relation* and *class of classes* in place of *class*.

George S. Boolos (1984) suggests a different reading, “There are some things, the *x*s, such that *y* is one of them.” Such a reading is available only in the second-order, one-place case, but there it seems to offer a way of avoiding overt quantification over classes or concepts. But it is controversial whether such plural quantification is prior to such notions as that of class, or whether the use of the plural involves a covert “ontological commitment” to something like classes. Boolos argues against the reduction of plural to class quantification, on the grounds that “[t]here are some classes such that any class is one of them iff it is not a member of itself” is true, while “[t]here is a class of classes such that any class” is false.

**GAME QUANTIFIERS.** Any first-order sentence is equivalent to one in *prenex* form, with all quantifiers out front. Any first-order prenex is equivalent to an existential second-order sentence (quantifying over functions from and to the domain *A* of the first-order variables), called its Skolem form, as with this equivalent pair (where the alternation of quantifiers may go on for any finite number *n* of rounds):

- (1)  $\forall x_1 \exists y_1 \forall x_2 \exists y_2 \dots \phi(x_1, y_1, x_2, y_2, \dots)$
- (2)  $\exists f_1 \exists f_2 \dots \forall x_1 \forall x_2 \dots \phi(x_1, f(x_1), x_2, f(x_1, x_2), \dots)$

Leon Henkin (1961) observes that one can associate to (1) a game for two players: player A chooses some  $a_1 \in A$ , player E chooses some  $b_1 \in A$ , A chooses  $a_2$ , then E chooses  $b_2, \dots$ , and in the end E wins if  $\phi[a_1, b_1, a_2, b_2, \dots]$ , and A if not. A strategy for a player is a rule telling that player how to play on each move as a function of the opponent’s previous moves. A winning strategy is one such that, if the player plays according to it, then the

player will win, regardless of how the opponent plays. A strategy for E can be represented as a pair of functions, one giving E’s first move as a function of A’s first move, the other giving E’s second move as a function of A’s first two moves. Then, (2) asserts that there is a winning strategy for E.

The game interpretation is especially useful if one wants to consider infinitely long formulas. A sentence like (1) but with an infinite alternation of quantifiers can be thought of as describing an infinite game—one may imagine each move made twice as fast as the one before—and the assertion that there exists a winning strategy for E is expressible as an infinitely long second-order sentence like (2) with infinite blocks of existential second-order and universal first-order quantifiers. There is this difference, that for a finite game one or the other of the players must have a winning strategy, but not for infinite games except in special cases. One such special case is that where  $\phi$  is a conjunction of formulas  $\phi_1, \phi_2, \dots$ , each involving only finitely many of the *x*’s and *y*’s. This game quantifier has a tractable theory in this case (see Moschovakis 1972).

**BRANCHING QUANTIFIERS.** Henkin (1961) also introduces branching quantifiers and suggests an interpretation in terms of an associated Skolem form, illustrated by the following pair:

- (3)  $\forall x_1 \exists y_1 \phi(x_1, y_1, x_2, y_2)$   
 $\forall x_2 \exists y_2$
- (4)  $\exists f_1 \exists f_2 \phi(x_1, f(x_1), x_2, f_2(x_2))$

Note the subtle difference between (4) and (2): In the latter,  $f_2$  is a one-place function. The main result about Henkin quantifiers is the Enderton-Walkoe theorem, asserting that not only is every Henkin quantifier sentence equivalent to an existential second-order sentence but also the converse holds. This means that known results about the logic of existential second-order sentences immediately apply to the logic of Henkin quantifier sentences: the Löwenheim-Skolem theorem, the compactness theorem, the definability of truth for sentences of this class by a sentence of the class, and more.

Jaako Hintikka (1996) introduces a nonbranching notation, in which (3) would be written as follows:

- (5)  $\forall x_1 \exists y_1 \forall x_2 \exists y_2 / x_1 \phi(x_1, y_1, x_2, y_2)$

The “/*x*<sub>1</sub>” is read “independent of *x*<sub>1</sub>.” Hintikka, long an advocate of a game interpretation of first-order quantifi-

cation, also suggests a game interpretation of the new quantifiers, in terms of a game of imperfect information, in which at the time of E's second move, E has available only information about A's second move, not about A's first move—which is most easily imagined if one thinks of E as a team, with different members making different moves and having available different information when doing so. Hintikka calls the logic with these quantifiers independence-friendly (or information-friendly) logic and makes strong and controversial claims about the philosophical significance of theorems about existential second-order sentences when restated for “IF” logic (see Hintikka 1996; compare Tennant 1998; see also Hodges 1997; Burgess 2003).

Which quantifiers considered by logicians have natural-natural language counterparts, and how close those counterparts are, is a much discussed question that cannot be addressed in this entry.

**See also** Artificial and Natural Languages; First-Order Logic; Intuitionism and Intuitionistic Logic; Quantifiers in Natural Language; Types, Theory of.

### **Bibliography**

- Barwise, K. Jon, and Solomon Feferman, eds. *Model-Theoretic Logics*. Berlin, Germany: Springer-Verlag, 1985.
- Bencivenga, Ermanno. “Free Logics.” In *Handbook of Philosophical Logic*. Vol. 3, *Alternatives to Classical Logic*, edited by Dov M. Gabbay and F. Guenther, 73–426. Dordrecht, Netherlands: D. Reidel, 1983.
- Benthem, J. F. A. K. van. “Tense Logic and Standard Logic.” *Logique et Analyse* 80 (1977): 41–83.
- Boolos, George S. “To Be Is to Be the Value of a Variable (or to Be Some Values of Some Variables).” *Journal of Philosophy* 81 (1984): 430–439.
- Burgess, John P. “A Remark on Henkin Sentences and Their Contraries.” *Notre Dame Journal of Formal Logic* 44 (2003): 185–188.
- Enderton, H. B. “Finite Partially-Ordered Quantifiers.” *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 16 (1970): 393–397.
- Henkin, Leon. “Some Remarks on Infinitely Long Formulas.” In *Infinitistic Methods: Proceedings of the Symposium on Foundations of Mathematics, Warsaw, 1959*, 167–183. Oxford, U.K.: Pergamon, 1961.
- Hintikka, Jaako. *The Principles of Mathematics Revisited*. New York: Cambridge University Press, 1996.
- Hodges, Wilfrid. “Compositional Semantics for a Language of Imperfect Information.” *Logic Journal of the IGPL* 5 (1997): 539–563.
- Kripke, Saul A. “Is There a Problem about Substitutional Quantification?” In *Essays in Semantics*, edited by Gareth Evans and John McDowell, 324–419. New York: Oxford University Press, 1976.
- Moschovakis, Yannis. “The Game Quantifier.” *Proceedings of the American Mathematical Society* 31 (1972): 245–250.
- Mostowski, Andrzej. “On a Generalization of Quantifiers.” *Fundamenta Mathematicæ* 44 (1957): 12–36.
- Quine, W. V. O. “Variables Explained Away.” *Proceedings of the American Philosophical Society* 104 (1960): 343–347.
- Tarski, Alfred. “What Are Logical Notions?” *History and Philosophy of Logic* 7 (1986): 143–154.
- Tennant, Neil. Review of *The Principles of Mathematics Revisited*, by Jaako Hintikka. *Philosophia Mathematica* 6 (1998): 90–115.
- Walkoe, W. J. “Finite Partially Ordered Quantification.” *Journal of Symbolic Logic* 35 (1970): 535–550.

*John P. Burgess (2005)*

## QUANTIFIERS IN NATURAL LANGUAGE

See Appendix, Vol. 10

## QUANTITY

See *Measurement and Measurement Theory*

## QUANTUM COMPUTING AND TELEPORTATION

In the 1980s and 1990s a series of revolutionary developments in the foundations of quantum mechanics led to what would later become the thriving fields of quantum information, quantum computation, and quantum cryptography. The roots of this revolution lie in the debate between Albert Einstein and Niels Bohr on the interpretation of quantum mechanics, specifically in the notion of “entangled” quantum states at the heart of the Einstein-Podolsky-Rosen argument for the incompleteness of quantum mechanics. What Einstein, Podolsky, and Rosen showed in their 1935 paper “Can Quantum-Mechanical Description of Physical Reality be Considered Complete?” was that composite quantum systems, consisting of spatially separated subsystems, could exist in certain states with peculiar nonclassical correlations between the outcomes of measurements on the subsystems. They argued that these correlations are incompatible with the assumption that the quantum state is a complete description of the system.

In a two-part commentary on the paper, Schrödinger referred to these states as being “entangled.” Roughly

thirty years later John Bell re-examined the Einstein-Podolsky-Rosen argument and showed that quantum mechanics could not be completed in the way Einstein would have liked, because the correlations of entangled states violate an inequality that an Einsteinian completion of quantum mechanics would have to satisfy. Essentially Bell showed that the correlations are inconsistent with any explanation in terms of a common cause (whether deterministic or stochastic) originating in the preparation of the state.

The salient feature of quantum information-processing tasks is the exploitation of entanglement as a new physical resource. Entanglement can be used to teleport quantum states, to exponentially outperform classical computers, and to implement cryptographic procedures that are impossible classically.

### TELEPORTATION AS REMOTE “STEERING”

Schroödinger regarded entangled states as problematic because they allow the possibility of what he called remote “steering,” which he regarded as unacceptable in a physical theory. As it turns out the teleportation of quantum states is an experimentally confirmed application of remote “steering” between two separated systems.

Consider Alice and Bob, the traditional protagonists in any two-party communication protocol. Suppose Alice and Bob each holds one of a pair of quantum particles associated with binary-valued physical quantities or “observables.” An example would be a pair of spin-½ particles, with two possible values, + and –, for the spin in some direction, say the z-direction. Alice’s particle might be represented by the pure quantum state  $|+\rangle_A$  and Bob’s particle by the pure quantum state  $|-\rangle_B$ . A spin state is represented as a unit vector in a 2-dimensional vector space, a so-called Hilbert space, the representation space for quantum states. The state of the composite two-particle system is a product state:

$$|+\rangle_A |-\rangle_B$$

represented by a vector in the 4-dimensional product Hilbert space for the two particles. An entangled pure state is a linear sum or “superposition” of product states that cannot itself be expressed as a product state. (More generally, for mixed states, representing mixtures or probability distributions of pure states, an entangled state is a state that cannot be represented as a convex combination or probability distribution of product states.)

Suppose Alice and Bob each holds one of a pair of particles in the entangled state:

$$|\psi\rangle = \frac{1}{\sqrt{2}} (|+\rangle_A |-\rangle_B - |-\rangle_A |+\rangle_B)$$

The coefficients (here  $\pm 1/\sqrt{2}$ ) can be complex numbers in general, and the squares of the absolute values of the coefficients (which are required to sum to 1: here ½ in both cases) represent the probabilities of obtaining the corresponding values of the relevant observables on measurement (+ and –, or – and +, for A and B). It turns out that Bob’s state, which defines the statistics for measurement outcomes on his particle, can be represented as an equal weight mixture of the orthogonal states  $|+\rangle_B$ ,  $|-\rangle_B$ , but equivalently as an infinity of other mixtures including, to take a specific example, the equal weight mixture of the four nonorthogonal states, represented as superpositions with complex coefficients  $\pm\alpha$ ,  $\pm\beta$  in the 2-dimensional Hilbert space of Bob’s particle:

$$|\phi_1\rangle_B = \alpha|+\rangle_B + \beta|-\rangle_B$$

$$|\phi_2\rangle_B = \alpha|+\rangle_B - \beta|-\rangle_B$$

$$|\phi_3\rangle_B = \beta|+\rangle_B + \alpha|-\rangle_B$$

$$|\phi_4\rangle_B = \beta|+\rangle_B - \alpha|-\rangle_B$$

If Alice measures the spin observable with outcomes associated with the two possible states  $|+\rangle_A$ ,  $|-\rangle_A$  on her particle A, and Bob measures the corresponding spin observable on his particle B, Alice’s outcomes will be oppositely correlated with Bob’s outcomes (+ with –, and – with +). If instead Alice prepares a spin-½ particle A’ in the state  $|\phi\rangle_{A'} = \alpha|+\rangle_{A'} + \beta|-\rangle_{A'}$  and measures an observable on the pair of systems A+A’ in her possession with possible outcomes corresponding to the four orthogonal states:

$$|1\rangle = \frac{1}{\sqrt{2}} (|+\rangle_{A'} |-\rangle_A - |-\rangle_{A'} |+\rangle_A)$$

$$|2\rangle = \frac{1}{\sqrt{2}} (|+\rangle_{A'} |-\rangle_A + |-\rangle_{A'} |+\rangle_A)$$

$$|3\rangle = \frac{1}{\sqrt{2}} (|+\rangle_{A'} |+\rangle_A - |-\rangle_{A'} |-\rangle_A)$$

$$|4\rangle = \frac{1}{\sqrt{2}} (|+\rangle_{A'} |+\rangle_A + |-\rangle_{A'} |-\rangle_A)$$

(the so-called Bell states), she will obtain the outcomes 1, 2, 3, 4 with equal probability, and these outcomes will be correlated with Bob’s states  $|\phi_1\rangle_B$ ,  $|\phi_2\rangle_B$ ,  $|\phi_3\rangle_B$ ,  $|\phi_4\rangle_B$  (i.e., if

Bob checks to see whether his particle is in the state  $|\phi_i\rangle_B$  when Alice reports that she obtained the outcome  $i=1, 2, 3, 4$ , he will find that this is always in fact the case). This follows because:

$$|\phi_1\rangle_{A'}|\psi\rangle = \frac{1}{2}(-|1\rangle|\phi_1\rangle_B - |2\rangle|\phi_2\rangle_B + |3\rangle|\phi_3\rangle_B + |4\rangle|\phi_4\rangle_B)$$

In this sense, Alice can “steer” Bob’s particle into any equivalent mixture generating the same statistics by an appropriate local measurement.

Now, remote “steering” in this probabilistic sense is precisely what makes quantum teleportation possible. Suppose Alice and Bob share a pair of spin- $\frac{1}{2}$  particles  $A$  and  $B$  in the entangled state and Alice is given a spin- $\frac{1}{2}$  particle  $A'$  in an *unknown* state  $|\phi_1\rangle$ . There is no procedure by which Alice can determine the unknown state, but if Alice measures the composite system  $A+A'$  in the Bell basis, she will “steer” Bob’s particle into one of the states  $|\phi_1\rangle_B, |\phi_2\rangle_B, |\phi_3\rangle_B, |\phi_4\rangle_B$  with equal probability. If Alice tells Bob the outcome of her measurement, Bob can apply a local operation corresponding to a transformation in the Hilbert space of his particle to obtain the state  $|\phi_1\rangle_B$ .

Note that before Alice sends Bob the outcome of her measurement, the quantum state that Bob assigns to his particle—the information represented by the mixed state—is unchanged by Alice’s measurement operation, even though after Alice’s measurement the probability is  $\frac{1}{4}$  that the state of Bob’s particle is in fact  $|\phi_1\rangle$  (in this case the local operation to obtain the state is represented by the identity). The trick that results in the transference of the state  $|\phi_1\rangle$  from Alice to Bob, without the particle  $A'$  traveling from Alice to Bob, is the ability afforded Alice by the shared entangled state to correlate one of four measurement outcomes (each occurring with probability  $\frac{1}{4}$ ) with one of four states that together represent a particular decomposition of Bob’s mixed state. The transference of the state of  $A'$  to Bob’s particle is accomplished by Bob’s operation, which requires that Alice sends the information about her measurement outcome to Bob. In the teleportation protocol the state of the particle  $A'$  is destroyed by Alice’s measurement and recreated as the state of Bob’s particle by Bob’s operation—in fact, the systems  $A$  and  $A'$  end up in an entangled state as the result of Alice’s measurement. (Note that if the state  $|\phi_1\rangle$  of  $A'$  were not destroyed there would be two copies of the state, which would violate the quantum “no cloning” theorem.)

### COMPUTATION VIA ENTANGLEMENT

The field of quantum computation was launched in the 1980s with two seminal papers by David Deutsch in 1985 and Richard Feynman in 1982. The basic idea can be

illustrated by the first genuinely quantum algorithm, proposed by Deutsch, later improved by Duetsch and Jozsa in 1992.

Consider a function  $f$  that maps an input value  $x=0$  or  $x=1$  onto an output value that is either 0 or 1. The algorithm for computing  $f$  might be quite complicated. To take Mermin’s example,  $f(x)$  might represent the value of the millionth bit in the binary expansion of  $\sqrt{2+x}$ , so that  $f(0)$  is the millionth bit in the expansion of  $\sqrt{2}$  while  $f(1)$  is the millionth bit in the expansion of  $\sqrt{3}$ . Suppose we are interested in whether the function  $f(x)$  is constant for both values of  $x$  or takes different values for both values of  $x$ —whether the millionth bit of  $\sqrt{2}$  is the same as the millionth bit of  $\sqrt{3}$ , or not. With a classical computer we would have to run through the algorithm twice to evaluate  $f(0)$  and  $f(1)$  and then compare these values. With a quantum computer it is possible to answer the question in a single run of the algorithm.

We might represent the computation of  $f$  by a classical computer as follows:

$$\langle 0 \rangle \langle 0 \rangle \rightarrow \langle 0 \rangle \langle f(0) \rangle$$

$$\langle 1 \rangle \langle 0 \rangle \rightarrow \langle 1 \rangle \langle f(1) \rangle$$

where  $\langle x \rangle \langle f(x) \rangle$  represents the input and output registers for the computation, and  $\rightarrow$  represents the mapping defined by the algorithm.

In the case of a quantum computer the input and output registers are quantum states, specifically here “qubits,” or states represented as orthogonal vectors in a 2-dimensional Hilbert space:

$$|0\rangle|0\rangle \rightarrow |0\rangle|f(0)\rangle$$

$$|1\rangle|0\rangle \rightarrow |1\rangle|f(1)\rangle$$

Here  $\rightarrow$  represents the quantum mechanical implementation of the algorithm by quantum transformations of the input state. We could put the input register into a superposition of quantum states  $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ , in which case,

since the quantum transformations are linear maps, the quantum implementation of the algorithm yields:

$$\frac{1}{\sqrt{2}} (|0\rangle + |1\rangle)|0\rangle \rightarrow \frac{1}{\sqrt{2}} (|0\rangle|f(0)\rangle + |1\rangle|f(1)\rangle)$$

This state is a linear superposition of *both* possible inputs and the associated outputs of  $f$ , apparently representing the computation for *all* possible values. Unfortunately however only a single read-out is possible: if we make a measurement on both registers, we obtain just one of the possible results with probability  $\frac{1}{2}$ , and the original state,

in which all possible inputs and associated outputs are represented, is altered. So there is no advantage over a classical computer.

Now we are interested in a global property of  $f$ , whether  $f$  is constant or balanced. Remarkably it turns out that we can answer this question in just one run of the algorithm, but at the expense of foregoing any information about the value of the function for either input. The final state is a product state (of the two registers) if  $f(0) = f(1)$  and an entangled state if  $f(0) \neq f(1)$ . By appropriate quantum transformations (see the discussion in Mermin's *Lecture Notes on Quantum Computation*) this state can be transformed to:

$$|1\rangle(|f(0)\rangle - |\overline{f(0)}\rangle)$$

if  $f(0) = f(1)$ , and to:

$$|0\rangle(|f(0)\rangle - |\overline{f(0)}\rangle)$$

if  $f(0) \neq f(1)$ , where  $\overline{0} = 1$  and  $\overline{1} = 0$ . The outcome of a measurement on the input register, + or -, will distinguish whether  $f(0) = f(1)$  or  $f(0) \neq f(1)$ .

Note that a measurement of the output register will yield the value  $f(0)$  or  $\overline{f(0)}$  with probability  $\frac{1}{2}$ , that is, 0 or 1 with probability  $\frac{1}{2}$ , which provides no information about the value of  $f$  for either input. In general a quantum computation involves the evolution of correlations in successive entangled states to a final state in which a measurement can determine the answer to a question about a global property of a function. The global property here is a disjunctive property:

$$(f(0) = f(1) = 0) \vee (f(0) = f(1) = 1)$$

or:

$$((f(0) = 1) \wedge (f(1) = 0)) \vee ((f(0) = 0) \wedge (f(1) = 1))$$

and the computation yields one or other of these disjunctions in the final measurement, which excludes the possibility of recording of the values of the disjuncts. The two alternative disjunctions are represented in the 4-dimensional Hilbert space of the two registers as quantum disjunctions, corresponding to two orthogonal 2-dimensional planes. Depending on whether the function is constant or balanced, the final state of the two registers is represented by a vector lying in one or the other of these two planes, and this can be determined by a measurement of the input register.

The Deutsch-Jozsa algorithm is a simple example of a quantum algorithm. More sophisticated quantum computation algorithms, such as Shor's algorithm for finding

the prime factors of a number, demonstrate an exponential speed-up over classical computation. The foundational significance of quantum computation concerns our understanding of computational complexity, that is, the relative efficiency of computational algorithms, rather than our characterization of the class of computable functions as those functions computable by a Turing machine. What quantum computation achieves is the possibility of solving a problem in a run-time (or number of computational steps) that increases as a polynomial function of the size of the input, while the computation using a classical computer would require superpolynomial, typically exponential, time. The difference can be quite dramatic. A classical computer of the sort available as of this writing would take an amount of time longer than the age of the universe to factor a 250-digit number into its two prime factors, using the fastest known algorithm. By contrast a quantum computer using Shor's algorithm could find the factors in minutes.

*See also* Bell, John, and Bell's Theorem; Bohr, Niels; Einstein, Albert; Hilbert, David; Many Worlds/Many Minds Interpretation of Quantum Mechanics; Quantum Mechanics; Schrödinger, Erwin.

### Bibliography

- Bell, John S. "On the Einstein-Podolsky-Rosen Paradox." *Physics* 1 (1964): 195–200. Reprinted in *Speakable and Unsayable in Quantum Mechanics*, John S. Bell. Cambridge, U.K.: Cambridge University Press, 1987.
- Bennett, Charles, Gilles Brassard, Claude Crepeau, Richard Jozsa, Asher Peres, and William Wootters. "Teleporting an Unknown Quantum State via Dual Classical and EPR Channels." *Physical Review Letters* 70 (1993): 1895–1899.
- Bohr, Niels. "Discussion with Einstein on Epistemological Problems in Modern Physics." In *Albert Einstein: Philosopher-Scientist*, edited by Paul A. Schilpp, 201–241. La Salle, IL: Open Court, 1949.
- Brassard, Gilles, Samuel Braunstein, and Richard Cleve. "Teleportation as a Quantum Computation." *Physica D* 120 (1998): 43–47.
- Brown, Julian. *Minds, Machines, and the Multiverse*. New York: Simon and Schuster, 2000.
- Deutsch, David. "Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer." *Proceedings of the Royal Society of London A* 400 (1985): 97.
- Deutsch, David, and Richard Jozsa. "Rapid Solution of Problems of Quantum Computation." *Proceedings of the Royal Society of London A* 425 (1968): 73–90.
- Einstein, Albert, Boris Podolsky, and Nathan Rosen. "Can Quantum-Mechanical Description of Physical Reality be Considered Complete?" *Physical Review* 47 (1935): 777–780.
- Feynman, Richard P. "Simulating Physics with Computers." *International Journal of Theoretical Physics* 21 (1982): 467–488.

- Mermin, David N. *Lecture Notes on Quantum Computation*. Available from <http://people.ccmr.cornell.edu/mermin/qcomp/CS483.html>.
- Nielsen, Michael, and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge: Cambridge University Press, 2000.
- Schrödinger, Erwin. "Discussion of Probability Relations Between Separated Systems." *Proceedings of the Cambridge Philosophical Society* 31 (1935): 555–563.
- Schrödinger, Erwin. "Probability Relations Between Separated Systems." *Proceedings of the Cambridge Philosophical Society* 32 (1936): 446–452.

*Jeffrey Bub (2005)*

## QUANTUM LOGIC AND PROBABILITY

Quantum physics predicts many astonishing physical effects that have been subsequently observed in the laboratory. Perhaps the most significant effect is the violation of Bell's inequality, which implies a failure of classical locality. But the most widely known bit of quantum magic is the experiment of Clinton Davisson and Lester Germer demonstrating interference effects for electrons. Richard Feynman said this is a phenomenon "which is impossible, *absolutely* impossible, to explain in any classical way, and which has in it the heart of quantum mechanics. In reality, it contains the *only* mystery" (1963–1965, Vol. 3, p. 1-1). As we will see, Feynman somewhat overstates the case, but let us first try to get in his frame of mind.

The interference effect is illustrated by the two-slit experiment. If we send a plain water wave toward a barrier with two narrow slits in it, we find that a circular wave is produced on the far side of each slit. As these two circular waves expand, they eventually overlap and interfere. Where the crest of one meets the crest of the other, we get a crest of twice the height; where the trough of one meets the trough of the other, we get a trough of twice the depth; and where the crest of one meets the trough of the other, the waves cancel out. This creates *interference bands*: regions of extreme agitation where the waves meet in phase, crest-to-crest and trough-to-trough, juxtaposed with quiescent regions where the waves meet out of phase, crest-to-trough. The pattern of regions of high and low activity is easy to calculate. Notice, in particular, that there are places where one would observe wave motion if either slit alone were open, but where there are no waves when both slits are open, because of destructive interference.

In quantum theory, a *wave function* represents the physical state of an electron, and for a *single* electron the wave function is mathematically similar to a water wave. (It is not exactly the same, since it is a *complex-valued* function. Moreover, this analogy works only for a *single* electron. The wave function for a pair of electrons is defined on the configuration space of the system, which has more dimensions than physical space.) The dynamics of the wave function is similar enough to the dynamics of water waves to display the same interference effects. That is, in the case of a single electron shot at a screen with a single slit, the wave function that makes it through the slit spreads out on the far side in a sort of circular pattern. And in the case of a single electron shot at a screen with two slits, the wave function that gets through spreads out in two circular patterns, one centered at each slit, and these interfere where they overlap, just like the water waves.

Of course, when we actually *look for* a single electron, we never find it spread out; we always find it at some localized place. We can use the wave function to make predictions about where the electron will appear by squaring the wave function and interpreting this value as the *probability* that the particle will be found at a particular location. If we do many identically prepared experiments, we find that the distribution of the electrons matches the square of the amplitude of the wave function, thereby confirming the predictive accuracy of quantum mechanics. But the mystery is this: To get these interference effects, we do not have to send many electrons *at the same time*. We can send the electrons through the device one at a time, with long gaps between them, and watch the interference bands build up slowly, dot by dot. So it is not that different electrons are somehow interfering *with each other*; it is rather that each electron is somehow interfering *with itself*.

To make the effect even more vivid, consider this fact. We send electrons through the slits one at a time and watch for flashes on a distant screen. There are particular areas on the distant screen where we will sometimes see flashes when only the right slit is open, and sometimes see flashes when only the left slit is open, but *never* see flashes (because of destructive interference) when both slits are open. So each electron sent through when both slits are open must somehow be physically influenced by the fact that both are open, since that region is only forbidden when both are open. But, to put the question fancifully, how can an electron "know" that both slits are open if (being a tiny particle) it only goes through one slit?

This would appear to be a straightforward *physical* question that calls for a *physical* answer. And indeed, two different physical answers are available, corresponding to the two straightforward ways to interpret the quantum formalism. On the one hand, if one thinks that the wave function is *complete*, that is, that it encodes all the physical characteristics of the electron, then one will simply reject that claim that the electron is a tiny particle that can only go through one slit. If the wave function is complete, then when the wave function spreads out enough to go through both slits, the electron itself spreads out enough to go through both slits, and that is how it can interfere with itself. This leaves a mystery, but the mystery is not why there is interference. Rather, it is why the electron makes a small, localized flash on the far screen. This problem is solved, in this approach, by giving an account of wave-function collapse.

The second physical answer maintains that the electron is indeed a tiny particle that always has a well-defined location, and hence goes through one slit or the other. On this approach, the wave function is *not* complete, since it does not indicate what that position is. This account is realized in the “pilot wave” theory of Louis de Broglie and David Bohm. As John Bell has written, “While the founding fathers agonized over the question: ‘particle’ or ‘wave’, de Broglie in 1925 proposed the obvious answer: ‘particle’ and ‘wave.’” That is, in the view of de Broglie and Bohm, there is, in addition to the located particle, a wave function that *guides the trajectory* of the particle. The state of the wave function is influenced by the fact that both slits are open, in exactly the way the quantum formalism indicates. So each particle “knows” that both slits are open, even though it goes through only one, because the *wave function* “knows” that both slits are open, and the wave function guides the particle.

These two physical answers to the puzzle of the interference bands are perfectly adequate, and evidently require no adjustments to classical logic or probability theory. The solution of de Broglie and Bohm is even, in certain sense, a classical solution, contrary to Feynman’s worry. So there is nothing in the phenomena discovered by modern physics that could *require* us to abandon or modify classical logic or probability theory.

Nonetheless, there have been many attempts, of various sorts, to argue that a change in logic or probability theory is at least *suggested* by the mathematical form of quantum theory, or that a change in logic or probability will produce an interpretation that is both physically adequate and somehow preferable to the two physical solutions outlined above.

At this point one would like a clear account of how classical logic or probability theory might be changed, and how the change might help us understand phenomena like the two-slit experiment without recourse to the sorts of physical hypotheses discussed above (hypotheses that, by the way, are already used to solve the measurement problem in quantum theory). Unfortunately, no such clear account is possible, because despite a long history and many attempts, no such account has ever been produced. So in its place, we must search instead for the reasons that anyone ever thought that classical logic or probability theory is responsible for the “mystery” surrounding these phenomena.

There are several different routes that can lead us to call into question classical logic. One, followed by Feynman in his famous *Lectures on Physics* (1963–1965), proceeds by reasoning about the two-slit experiment. The other, which is the foundation of the technical field of quantum logic, proceeds from an analysis of the mathematical machinery of quantum theory. Let us examine these in turn.

In his analysis of the two-slit experiment, Feynman first introduced proposition A:

*Proposition A.* Each electron *either* goes through slit 1 *or* goes through slit 2.

Feynman then went on to consider what he calls the *consequences* of this proposition for predictions about the results of the experiment. If proposition A is true, he said, then we ought to be able to calculate the probability that the electron will land at any point of the screen by first determining the probability for electrons that go through slit 1 (by blocking slit 2 and seeing what happens), and then determining the probability for electrons that pass through slit 2 (by blocking slit 1 and seeing what happens). If proposition A is true when both slits are open, Feynman said, then the individual probabilities derived from these experiments should add. With both slits open, there are more ways for any result to come about (since an electron can get to a certain spot either by going through slit 1 or by going through slit 2), and the chance of the result should be just the sum of the chances of each process. This is, of course, not what we see. Because of the interference, there are places on the far screen where electrons appear with either slit open, but where no electrons appear with both slits open. Feynman concluded, “When one does *not* try to tell which way the electron goes, when there is nothing in the experiment to disturb the electrons, then one may *not* say that an electron goes through either hole 1 or hole 2. If one does say that, and starts to make deductions from the statement, he will make errors

in the analysis.” That is, Feynman concludes from considerations of how the probabilities ought to add that proposition A is not true.

We can equally well present Feynman’s dilemma using only logic rather than probability theory. There are places on the screen where an electron can appear when only slit 1 is open (and the electron goes through slit 1) and also when only slit 2 is open (and the electron goes through slit 2). So if the electron goes through slit 1, it can appear at a certain point, and if it goes through slit 2, it can appear at that same point. From the premise that the electron either goes through slit 1 or goes through slit 2, it then follows by a disjunctive syllogism that it can appear at that point. But with *both* slits open, the electron cannot appear at that point. It seems to follow that when both slits are open, the disjunction is not true. It is not the case that the electron went through slit 1 or went through slit 2.

Something must have gone wrong with Feynman’s analysis somewhere. For in the theory of de Broglie and Bohm, electrons always have exact locations, and every electron that gets from the source to the far screen goes either through slit 1 or through slit 2. And the de Broglie and Bohm theory makes all the right predictions: exactly the predictions of quantum theory. Where did Feynman go wrong?

The solution is not hard to seek. Feynman considers first doing an experiment with slit 1 open *and* slit 2 closed, and then an experiment with slit 2 open *and* slit 1 closed. So the experimentally confirmed propositions are that if the electron goes through slit 1 with slit 2 closed, it can appear at a certain spot, and that if it goes through slit 2 with slit 1 closed, it can appear at that spot. The relevant disjunction for using disjunctive syllogism is the following: The particle either goes through slit 1 with slit 2 closed or through slit 2 with slit 1 closed. From this disjunction it does indeed follow that the electron can appear at the spot. But this disjunction tells us nothing at all about what can happen *with both slits open*.

Feynman’s thought, evidently, is that if the electron goes through one slit, then it cannot make any difference whether the other slit is open. This is a reasonable conjecture, supported by classical intuitions. But this conjecture is false, and quantum theory shows why it is false: The state of the wave function is influenced by the state of both slits. Indeed, one consequence of quantum mechanics is that the state of the wave function is influenced by the presence or absence of detectors at either slit. Even when both slits are open, a detector at one slit will cause the interference to go away *even when the detector does not*

*fire*. This is a straightforward mathematical consequence of the dynamics of the wave function. The ultimate *physical* moral is that one must take account of the entire experimental arrangement when considering what quantum mechanics predicts. As John Bell put it, “When one forgets the role of the apparatus . . . , one despairs of ordinary logic. . . . Hence ‘quantum logic.’ When one remembers the role of the apparatus, ordinary logic is just fine.” And the apparatus in question is the *whole experimental situation*, including elements (such as the presence or absence of detectors that *do not fire*) that would be deemed irrelevant in classical physics.

Feynman’s argument is a *physical* argument: It proceeds solely from the observation of experimental results to the (incorrect) conclusion that proposition A cannot be true, since one could deduce a false consequence from it. The field of quantum logic takes the opposite tack. Quantum logicians want to maintain that something like proposition A is *true* when both slits are open. But since false claims can apparently be deduced from proposition A using classical logic, this requires a change in logic itself.

Quantum logicians tend not to start from experiment, as Feynman does, but from observations about the form of the mathematical apparatus used in quantum theory. In particular, they begin with the observation that the space of all wave functions is a complex *vector space*. This means that given any pair of wave functions  $|\psi_1\rangle$  and  $|\psi_2\rangle$ , and any two complex numbers  $\alpha$  and  $\beta$ , there exists another wave function of the form  $\alpha|\psi_1\rangle + \beta|\psi_2\rangle$ . Such a wave function is called a *superposition* of  $|\psi_1\rangle$  and  $|\psi_2\rangle$ .

Suppose that the wave function of an electron that goes through slit 1 with slit 2 closed is  $|\psi_1\rangle$ , and the wave function of an electron that goes through slit 2 with slit 1 closed is  $|\psi_2\rangle$ . Then when *both* slits are open, the wave function will be of the form  $\alpha|\psi_1\rangle + \beta|\psi_2\rangle$ , a superposition of  $|\psi_1\rangle$  and  $|\psi_2\rangle$ . (In particular, in the usual experimental configuration, it will be  $(1/\sqrt{2})|\psi_1\rangle + (1/\sqrt{2})|\psi_2\rangle$ .) This wave function is evidently neither  $|\psi_1\rangle$  nor  $|\psi_2\rangle$ . It would *not* be correct to say, with classical logical connectives, that the electron is either in state  $|\psi_1\rangle$  or in state  $|\psi_2\rangle$ . So if we allow  $|\psi_1\rangle$  now to stand for the *proposition* that the electron is in state  $|\psi_1\rangle$ , and  $|\psi_2\rangle$  to stand for the proposition that the electron is in state  $|\psi_2\rangle$ , then the classical proposition “ $|\psi_1\rangle$  or  $|\psi_2\rangle$ ” is *false* when both slits are open.

What the quantum logician does, though, is to introduce a *new* connective, usually written  $\vee$ , that is defined so that  $|\psi_1\rangle \vee |\psi_2\rangle$  is true whenever the electron is in a superposition of  $|\psi_1\rangle$  and  $|\psi_2\rangle$ . If one tries to think of  $\vee$  as



a sort of disjunction, one can then have a disjunction that is true even though neither disjunct is true—a circumstance that violates classical truth conditions.

More technically, the quantum logician associates propositions with subspaces of Hilbert space, the vector space of the wave function. The “conjunction” of two propositions (written  $\mathbf{A} \wedge \mathbf{B}$ ) is just the *intersection* of the associated subspaces, and the “disjunction” of two propositions ( $\mathbf{A} \vee \mathbf{B}$ ) is the *span* of the subspaces, that is, the subspace consisting of all vectors that can be formed by adding vectors from the two given subspaces. A proposition is *true* just in case the wave function of the system lies in the associated subspace. So if the wave function of the system is  $(1/\sqrt{2})|\psi_1\rangle + (1/\sqrt{2})|\psi_2\rangle$ , then the proposition  $|\psi_1\rangle$  is not true, and the proposition  $|\psi_2\rangle$  is not true, but the proposition  $|\psi_1\rangle \vee |\psi_2\rangle$  is true.

With this terminology in place, one can easily show that the set of “quantum propositions” form a non-Boolean (nondistributive) lattice under the operations  $\vee$  and  $\wedge$ . This is a straightforward mathematical fact about the structure of subspaces of Hilbert space under these operations. There is no nonclassical logic or probability theory here, just standard mathematics.

Of course, if one starts to pronounce  $\vee$  “or” and  $\wedge$  “and,” then matters can get somewhat confusing. Because the lattice of quantum propositions is nondistributive,  $(\mathbf{A} \vee \mathbf{B}) \wedge \mathbf{C}$  can be true while  $(\mathbf{A} \wedge \mathbf{C}) \vee (\mathbf{B} \wedge \mathbf{C})$  is false. If one presents this fact by saying that “(A or B) and C” is true while “(A and C) or (B and C)” is false, then it appears that de Morgan’s laws have failed. Hence the supposed need for quantum logic.

If quantum logic is just the study of the structure of subspaces of Hilbert space, then it is a perfectly legitimate, but badly named, enterprise. It is not an alternative to, or replacement for, classical logic, since it studies connectives that are not the classical connectives. Nothing has been shown to be wrong or misleading about classical logic. Rather, the problem lies with our intuitions about experimental conditions, which lead us incorrectly to expect that whether the second slit is open is irrelevant to the behavior of the electron at the other slit. Quantum mechanics shows not that there is anything wrong with classical logic, but rather that the physics of the quantum world is very unlike the physics of Isaac Newton and James Clerk Maxwell. The surprising relevance of experimental conditions is shown by experiments like the two-slit experiment, and the appropriate way to reason about these experiments is, of course, classically.

What about Feynman’s proposition A? With both slits open, is it correct or incorrect to say that the electron either went through slit 1 or slit 2? The answer to this question once again depends on physics rather than logic. If the de Broglie and Bohm theory is correct, then the electron always goes through one slit or the other. Retrospectively, one can even tell which slit it went through. Proposition A is therefore true. If one adopts an interpretation according to which the wave function is complete, then the wave function is all there is to the electron, and the wave function “goes through” both slits. Part of it goes through each slit, so it goes neither *entirely* through slit 1 nor *entirely* through slit 2. On a truth-functional reading of “or,” proposition A is false. As long one is clear about the exact content of any proposition and about the interpretation of quantum theory at issue, classical logic and probability theory work just fine.

*See also* Bell, John, and Bell’s Theorem; Bohm, David; Hilbert, David; Logic, Non-Classical; Maxwell, James Clerk; Newton, Isaac; Non-locality; Quantum Mechanics.

### Bibliography

- Birkhoff, Garrett, and John von Neumann. “The Logic of Quantum Mechanics.” *Annals of Mathematics* 37 (1936): 823–843.
- Feynman, Richard, Robert B. Leighton, and Matthew Sands. *The Feynman Lectures on Physics*. Reading, MA: Addison-Wesley, 1963–1965.
- Kochen, Simon, and Ernst P. Specker. “The Problem of Hidden Variables in Quantum Mechanics.” *Journal of Mathematics and Mechanics* 17 (1967): 59–87.
- Maudlin, Tim. “The Tale of Quantum Logic.” In *The Philosophy of Hilary Putnam*, edited by Yemima Ben Menahem. New York: Cambridge University Press, 2005.
- Putnam, Hilary. “Is Logic Empirical?” In *Boston Studies in the Philosophy of Science*. Vol. 5, edited by Robert S. Cohen and Marx Wartofsky. Dordrecht, Netherlands: D. Reidel, 1968. Reprinted as “The Logic of Quantum Mechanics,” in his *Mathematics, Matter, and Method*. New York: Cambridge University Press, 1976.
- Van Fraassen, Bas. *Quantum Mechanics: An Empiricist View*. New York: Oxford University Press, 1992.
- Von Neumann, John. *Mathematische Grundlagen der Quantenmechanik*. Berlin: Springer-Verlag, 1932. Translated by Robert T. Beyer as *Mathematical Foundations of Quantum Mechanics*. Princeton, NJ: Princeton University Press, 1955.

**Tim Maudlin (2005)**

## QUANTUM MECHANICS

Quantum mechanics has the distinction of being considered both the most empirically successful and the most poorly understood theory in the history of physics.

To take an oft-cited example of the first point: The theoretically calculated value of the anomalous magnetic moment of the electron using quantum electrodynamics matches the observed value to twelve decimal places, arguably the best confirmed empirical prediction ever made. To illustrate the second point, we have the equally oft-cited remarks of Niels Bohr, “Anyone who says that they can contemplate quantum mechanics without becoming dizzy has not understood the concept in the least,” and of Richard Feynman, “[We] have always had (secret, secret, close the doors!) we always have had a great deal of difficulty in understanding the world view that quantum mechanics represents.” How could both of these circumstances obtain?

For the purposes of making predictions, quantum theory consists in a mathematical apparatus and has clear enough rules of thumb about how to apply the mathematical apparatus in various experimental situations. If one is doing an experiment or observing something, one must first associate a mathematical *quantum state* or *wave function* with the system under observation. For example, if one prepares in the laboratory an electron beam with a fixed momentum, then the quantum state of each electron in the beam will be something like a sine wave. In the case of a single particle it is common to visualize this wave function as one would a water wave: as an object extended in space. Although this visualization works for a single particle, it does not work in general, so care must be taken. But for the moment, this simple visualization works. The wave function for the electron is “spread out” in space.

The second part of the mathematical apparatus is a *dynamical equation* that specifies how the quantum state changes with time so long as no observation or measurement is made on the system. These equations have names like the *Schrödinger equation* (for nonrelativistic quantum mechanics) and the *Dirac equation* (for relativistic quantum field theory). In the case of the electron mentioned earlier the dynamical equation is relevantly similar to the dynamical equation for water waves, so we can visualize the quantum state as a little plane water wave moving in a certain direction. If the electron is shot at a screen with two slits in it, then the quantum state will behave similarly to a water wave that hits such a barrier: circularly expanding waves will emerge from each slit,

and there will be constructive and destructive interference where those waves overlap. If beyond the slits there is a fluorescent screen, we can easily calculate what the quantum state “at the screen” will look like: It will have the peaks and troughs characteristic of interfering water waves.

Finally comes the interaction with the screen. Here is where things get tricky. One would naively expect that the correct way to understand what happens when the electron wave function reaches the screen is to build a physical model of the screen and apply quantum mechanics to it. But that is not what is done. Instead, the screen is treated as a measuring device and the interaction with the screen as a measurement, and new rules are brought into play.

The new rules require that one first decide what property the measuring device measures. In the case of a fixed screen it is taken that the screen measures the *position* of a particle. If instead of a fixed screen we had an absorber on springs, whose recoil is recorded, then the device would measure the *momentum* of the particle. These determinations are typically made by relying on classical judgments: There is no algorithm for determining what a generic (physically specified) object “measures,” or indeed whether it measures anything at all. But laboratory apparatus for measuring position and momentum have been familiar from before the advent of quantum theory, so this poses no real practical problem.

Next, the property measured gets associated with a mathematical object called a *Hermitian operator*. Again, there is no algorithm for this, but for familiar classical properties like position and momentum the association is established. For each Hermitian operator there is an associated set of wave functions called the *eigenstates* of the operator. It is purely a matter of mathematics to determine the eigenstates. Each eigenstate has associated with it an *eigenvalue*: The eigenvalues are supposed to correspond to the possible outcomes of a measurement of the associated property, such as the possible values of position, momentum, or energy. (Conversely, it is typically assumed that for every Hermitian operator, there corresponds a measurable property and possible laboratory operations that would measure it, although there is no general method for specifying these.)

The last step in the recipe for making predictions can now be taken. When a system is measured, the wave function for the system is first expressed as a sum of terms, each term being an eigenstate of the relevant Hermitian operator. Any wave function can be expressed as a sum of such terms, with each term given a weight, which is a

complex number. For example, if an operator has only two eigenstates, call them  $|1\rangle$  and  $|2\rangle$ , then any wave function can be expressed in the form  $\alpha|1\rangle + \beta|2\rangle$ , with  $\alpha$  and  $\beta$  complex numbers such that  $|\alpha|^2 + |\beta|^2 = 1$ . (This is the case, for example, when we measure the so-called spin of an electron in a given direction, and always get one of two results: spin up or spin down.) Recall that each eigenstate is associated with a possible outcome of the measurement:  $|1\rangle$ , for example, could be associated with getting spin up, and  $|2\rangle$  with getting spin down. The quantum mechanical prediction is now typically a probabilistic one: the chance of getting the result associated with  $|1\rangle$  is  $|\alpha|^2$ , and the chance of getting the result associated with  $|2\rangle$  is  $|\beta|^2$ . In general, one writes out the wave function of the system in terms of the appropriate eigenstates, and then the chance of getting the result associated with some eigenstate is just the square of the complex number that weights the state.

We can now see how quantum theory makes empirical predictions: So long as one knows the initial quantum state of the system and the right Hermitian operator to associate with the measurement, the theory will allow one to make probabilistic predictions for the outcome. Those predictions turn out to be exquisitely accurate.

If a Hermitian operator has only a finite number of eigenstates, or the eigenvalues of the operator are discrete, then any associated measurement should have only a discrete set of possible outcomes. This has already been in the case of spin; for a spin-1/2 particle such as an electron, there are only two eigenstates for the spin in a given direction. Physically, this means that when we do an experiment to measure spin (which may involve shooting a particle through an inhomogeneous magnetic field) we will get only one of two results: Either the particle will be deflected up a given amount or down a given amount (hence spin up and spin down). In this case the physical quantity is *quantized*; it takes only a discrete set of values. But quantum theory does not require all physical magnitudes to be quantized in this way; the position, momentum, or energy of a free particle is not. So the heart of quantum theory is not a theory of discreteness, it is rather just the mathematical apparatus and the rules of application described earlier.

## THE MEASUREMENT PROBLEM

Why, then, is the quantum theory so puzzling, or so much more obscure than, say, classical mechanics? One way that it differs from classical theory is that it provides only probabilistic predictions for experiments, and one might well wonder, as Albert Einstein famously did, whether

this is because “God plays dice with the universe” (i.e., the physical world itself is not deterministic) or whether the probabilities merely reflect our incomplete knowledge of physical situation. But even apart from the probabilities, the formulation of the theory is rather peculiar. Rules are given for representing the physical state of a system and for how that physical state evolves and interacts with other systems when no measurement takes place. This evolution is perfectly deterministic. A different set of rules is applied to derive predictions for the outcomes of experiments, and these rules are not deterministic. Still, an experiment in a laboratory is just a species of physical interaction, and ought to be treatable as such. There should be a way to describe the physical situation in the lab, and the interaction of the measured system with the measuring device, that relies only on applying, say, the Schrödinger equation to the physical state of the system plus the lab.

John S. Bell put this point succinctly, “If you make axioms, rather than definitions and theorems, about the ‘measurement’ of anything else, then you commit redundancy and risk inconsistency” (1987, p. 166). You commit redundancy because while the axioms about measurement specify what should happen in a measurement situation, the measurement situation, considered as a simple physical interaction, ought also to be covered by the general theory of such interactions. You risk inconsistency because the redundancy produces the possibility that the measurement axioms will contradict the results of the second sort of treatment. This is indeed what happens in the standard approaches to quantum mechanics. The result is called the *measurement problem*.

The measurement problem arises from a conflict in the standard approach between treating a laboratory operation as a normal physical interaction and treating it as a measurement. To display this conflict, we need some way to represent the laboratory apparatus as a physical device and the interaction between the device and the system as a physical interaction. Now this might seem to be a daunting task; a piece of laboratory apparatus is typically large and complicated, comprising astronomically large numbers of atoms. By contrast, exact wave functions are hard to come by for anything much more complicated than a single hydrogen atom. How can we hope to treat the laboratory operation at a fundamental level?

Fortunately, there is a way around this problem. Although we cannot write down, in detail, the physical state of a large piece of apparatus, there are conditions that we must assume if we are to regard the apparatus as a good measuring device. There are necessary conditions

for being a good measuring device, and since we do regard certain apparatus as such devices, we must be assuming that they meet these conditions.

Take the case of spin. If we choose a direction in space, call it the  $x$ -direction, then there is a Hermitian operator that gets associated with the quantity  $x$ -spin. That operator has two eigenstates, which we can represent as  $|x\text{-up}\rangle_s$  and  $|x\text{-down}\rangle_s$ . The subscript  $s$  indicates that these are states of the system to be measured. We have pieces of laboratory equipment that can be regarded as good devices for measuring the  $x$ -spin of a particle. We can prepare such an apparatus in a state, call it the “ready” state, in which it will function as a good measuring device. Again, we do not know the exact physical details of this ready state, but we must assume such states exist and can be prepared. What physical characteristics must such a ready state have?

Besides the ready state, the apparatus must have two distinct indicator states, one of which corresponds to getting an “up” result of the measurement and the other that corresponds to getting a “down” result. And the key point about the physics of the apparatus is this: It must be that if the device in its ready state interacts with a particle in the state  $|x\text{-up}\rangle_s$ , it will evolve into the indicator state that is associated with the up result, and if it interacts with a particle in state  $|x\text{-down}\rangle_s$ , it will evolve into the other indicator state.

This can be put in a formal notation. The ready state of the apparatus can be represented by  $|\text{ready}\rangle_A$ , the up indicator state by  $|\text{“up”}\rangle_A$ , and the down indicator state by  $|\text{“down”}\rangle_A$ . If we feed an  $x$ -spin up particle into the device, the initial physical state of the system plus apparatus is represented by  $|x\text{-up}\rangle_s|\text{ready}\rangle_A$ , if we feed in an  $x$ -spin down particle the initial state is  $|x\text{-down}\rangle_s|\text{ready}\rangle_A$ . If the apparatus is, in fact, a good  $x$ -spin measuring device, then the first initial state must evolve into a state in which the apparatus indicates up, that is, it must evolve into  $|x\text{-up}\rangle_s|\text{“up”}\rangle_A$ , and the second initial state must evolve into a state that indicates down, that is,  $|x\text{-down}\rangle_s|\text{“down”}\rangle_A$ . Using an arrow to represent the relevant time evolution, then, we have for any good  $x$ -spin measuring device

$$|x\text{-up}\rangle_s|\text{ready}\rangle_A \rightarrow |x\text{-up}\rangle_s|\text{“up”}\rangle_A \text{ and}$$

$$|x\text{-down}\rangle_s|\text{ready}\rangle_A \rightarrow |x\text{-down}\rangle_s|\text{“down”}\rangle_A.$$

We have not done any real physics yet, we have just indicated how the physics must come out if there are to be items that count as good  $x$ -spin measuring devices, as we think there are.

The important part of the physics that generates the measurement problem is the arrow in the representations listed earlier, the physical evolution that takes one from the initial state of the system plus apparatus to the final state. Quantum theory provides laws of evolution for quantum states such as the Schrödinger and Dirac equations. These would be the equations one would use to model the evolution of the system plus apparatus as a normal physical evolution. And all these dynamical equations have a common mathematical feature; they are all linear equations. It is this feature of the quantum theory that generates the measurement problem, so we should pause over the notion of linearity.

The set of wave functions used in quantum theory form a *vector space*. This means that one can take a weighted sum of any set of wave functions and get another wave function. (The weights in this case are complex numbers, hence it is a complex vector space.) This property was mentioned earlier when it was noted that any wave function can be expressed as a weighted sum of the eigenvectors of an observable. An operator on a vector space is just an object that maps a vector as input to another vector as output. If the operator  $O$  maps the vector  $\mathbf{A}$  to the vector  $\mathbf{B}$ , we can write that as

$$O(\mathbf{A}) = \mathbf{B}.$$

A linear operator has the feature that you get the same result whether to operate on a sum of two vectors or you first operate on the vectors and then takes the sum. That is, if  $O$  is a linear operator, then for all vectors  $\mathbf{A}$  and  $\mathbf{B}$ ,

$$O(\mathbf{A} + \mathbf{B}) = O(\mathbf{A}) + O(\mathbf{B}).$$

The dynamical equations evidently correspond to operators; they take as input the initial physical state and give as output the final state, after a specified period has elapsed. But further, the Schrödinger and Dirac equations correspond to linear operators. Why is this important?

We have already seen how the physical state of a good  $x$ -spin measuring device must evolve when fed a particle in the state  $|x\text{-up}\rangle_s$  or the state  $|x\text{-down}\rangle_s$ . But these are not the only spin states that the incoming particle can occupy. There is an infinitude of spin states, which correspond to all the wave functions that can be expressed as  $\alpha|x\text{-up}\rangle_s + \beta|x\text{-down}\rangle_s$ , with  $\alpha$  and  $\beta$  complex numbers such that  $|\alpha|^2 + |\beta|^2 = 1$ . Correspondingly, there is an infinitude of possible directions in space in which one can orient a spin measuring device, and each of the directions is associated with a different Hermitian operator. For a direction at right angles to the  $x$ -direction, call it the  $y$ -direction, there are eigenstates  $|y\text{-up}\rangle_s$  and

$|y\text{-down}\rangle_s$ . These states can be expressed as weighted sums of the  $x$ -spin eigenstates, and in the usual notation

$$|y\text{-up}\rangle_s = 1/\sqrt{2}|x\text{-up}\rangle_s + 1/\sqrt{2}|x\text{-down}\rangle_s \text{ and}$$

$$|y\text{-down}\rangle_s = 1/\sqrt{2}|x\text{-up}\rangle_s - 1/\sqrt{2}|x\text{-down}\rangle_s.$$

So what happens if we feed a particle in the state  $|y\text{-up}\rangle_s$  into the good  $x$ -spin measuring device?

Empirically, we know what happens: About half the time the apparatus ends up indicating “up” and about half the time it ends up indicating “down.” There is nothing we are able to do to control the outcome:  $y$ -up eigenstate particles that are identically prepared nonetheless yield different outcomes in this experiment.

If we use the usual predictive apparatus, we also get this result. The “up” result from the apparatus is associated with the eigenstate  $|x\text{-up}\rangle_s$  and the “down” result associated with  $|x\text{-down}\rangle_s$ . The general recipe tells us to express the incoming particle in terms of these eigenstates as  $1/\sqrt{2}|x\text{-up}\rangle_s + 1/\sqrt{2}|x\text{-down}\rangle_s$ , and then to take the squares of the weighting factors to get the probabilities of the results. This yields a probabilistic prediction of 50 percent chance “up” and 50 percent chance “down,” which corresponds to what we see in the lab.

But if instead of the usual predictive apparatus we use the general account of physical interactions, we get into trouble. In that case, we would represent the initial state of the system plus apparatus as  $|y\text{-up}\rangle_s|ready\rangle_A$ . The dynamical equation can now be used to determine the physical state of the system plus apparatus at the end of the experiment.

But the linearity of the dynamical equations already determines what the answer must be. For

$$|y\text{-up}\rangle_s|ready\rangle_A = (1/\sqrt{2}|x\text{-up}\rangle_s + 1/\sqrt{2}|x\text{-down}\rangle_s)|ready\rangle_A$$

$$= 1/\sqrt{2}|x\text{-up}\rangle_s|ready\rangle_A + 1/\sqrt{2}|x\text{-down}\rangle_s|ready\rangle_A.$$

But we know how each of the two terms of this superposition must evolve, since the apparatus is a good  $x$ -spin measuring device. By linearity, this initial state must evolve into the final state

$$1/\sqrt{2}|x\text{-up}\rangle_s|“up”\rangle_A + 1/\sqrt{2}|x\text{-down}\rangle_s|“down”\rangle_A.$$

That is, the final state of the apparatus plus system must be a superposition of a state in which the apparatus yields the result “up” and a state in which the apparatus yields the result “down.” That is what treating the measurement as a normal physical interaction must imply.

So by making axioms about measurements, we have both committed redundancy and achieved inconsistency. The axioms say that the outcome of the experiment is not determined by the initial state; each of two outcomes is possible, with a 50 percent chance of each. But the treatment of the measurement as a normal physical interaction implies that only one final physical state can occur. And furthermore, that final physical state is an extremely difficult one to understand. It appears to be neither a state in which the measuring apparatus is indicating “up” nor a state in which the apparatus is indicating “down,” but some sort of symmetric combination of the two. If all the physical facts about the apparatus are somehow represented in its wave function, then it seems that at the end of the experiment the apparatus can neither be indicating up (and not down) nor down (and not up). But we always see one or the other when we do this experiment.

At this point our attention must clearly be turned to the mathematical object we have called the wave function. The wave function is supposed to represent the physical state of a system. The question is whether the wave function represents all of the physical features of a system, or whether systems represented by the same wave function could nevertheless be physically different. If one asserts the former, then one believes that the wave function is complete, if the latter, then the wave function is incomplete. The standard interpretations of the quantum formalism take the wave function to be complete, interpretations that take it to be incomplete are commonly called *hidden variables* theories (although that is a misleading name).

The wave function  $1/\sqrt{2}|x\text{-up}\rangle_s|“up”\rangle_A + 1/\sqrt{2}|x\text{-down}\rangle_s|“down”\rangle_A$  does not represent the apparatus as indicating up (and not down) or as indicating down (and not up). So if the wave function is complete, the apparatus, at the end of the experiment, must neither be indicating up (and not down) nor down (and not up). But that flatly contradicts our direct experience of such apparatus. This is the measurement problem. As Bell puts it, “Either the wave function, as given by the Schrödinger equation, is not everything, or it is not right” (1987, p. 201).

## COLLAPSE INTERPRETATIONS

**COLLAPSE TIED TO OBSERVATION.** What is one to do? From the beginning of discussions of these matters, Einstein held the argument to show that the wave function is not everything and hence that quantum mechanics is incomplete. The wave function might represent part of the physical state of a system, or the wave function might

represent some features of ensembles, collections, or systems, but the wave function cannot be a complete representation of the physical state of an individual system, like the particular  $x$ -spin measuring device in the laboratory after a particular experiment is done. For after the experiment, the apparatus evidently either indicates “up” or it indicates “down,” but the wave function does not represent it as doing so.

By contrast, the founders of the quantum theory, especially Bohr, insisted that the wave function is complete. And they did not want to deny that the measuring device ends up indicating one determinate outcome. So the only option left was to deny that the wave function, as given by the Schrödinger equation, is right. At some times, the wave function must evolve in a way that is not correctly described by the Schrödinger equation. The wave function must “collapse.” The standard interpretation of quantum mechanics holds that the wave function evolves, at different times, in either of two different ways. This view was given its canonical formulation in John von Neumann’s *Mathematical Foundations of Quantum Mechanics* (1955). Von Neumann believed (incorrectly, as we will see) that he had proven the impossibility of supplementing the wave function with hidden variables, so he thought the wave function must be complete. When he comes to discuss the time evolution of systems, Von Neumann says “[w]e therefore have two fundamentally different types of interventions which can occur in a system  $S$ . ... First, the arbitrary [i.e., nondeterministic] changes by measurement. ... Second, the automatic [i.e., deterministic] changes which occur with the passage of time” (p. 351). The second type of change is described by, for example, the Schrödinger equation, and the first by an indeterministic process of collapse.

What the collapse dynamics must be can be read off from the results we want together with the thesis that the wave function is complete. For example, in the  $x$ -spin measurement of the  $y$ -spin up electron, we want there to be a 50 percent chance that the apparatus indicates “up” and a 50 percent chance that it indicates “down.” But the only wave function that represents an apparatus indicating “up” is  $|\text{“up”}\rangle_A$ , and the only wave function for an apparatus indicating “down” is  $|\text{“down”}\rangle_A$ . So instead of a deterministic transition to the final state

$$1/\sqrt{2}|x\text{-up}\rangle_S|\text{“up”}\rangle_A + 1/\sqrt{2}|x\text{-down}\rangle_S|\text{“down”}\rangle_A$$

we must postulate an indeterministic transition with a 50 percent chance of yielding  $|x\text{-up}\rangle_S|\text{“up”}\rangle_A$  and a 50 percent chance of yielding  $|x\text{-down}\rangle_S|\text{“down”}\rangle_A$ .

It is clear what the collapse dynamics must do. What is completely unclear, though, is when it must do it. All Von Neumann’s rules say is that we get collapses when measurements occur and deterministic evolutions “with the passage of time.” But surely measurements also involve the passage of time; so under exactly what conditions do each of the evolutions obtain? Collapse theories, which postulate two distinct and incompatible forms of evolution of the wave function, require some account of when each type of evolution occurs.

Historically, this line of inquiry was influenced by the association of the problem with “measurement” or “observation.” If one begins with the thought that the non-linear evolution happens only when a measurement or observation occurs, then the problem becomes one of specifying when a measurement or observation occurs. And this in turn suggests that we need a characterization of an observer who makes the observation. Pushing even further, one can arrive at the notion that observations require a conscious observer of a certain kind, folding the problem of consciousness into the mix. As Bell asks, “What exactly qualifies some physical systems to play the role of ‘measurer’? Was the wave function of the world waiting to jump for thousands of millions of years until a single-celled living creature appeared? Or did it have to wait a little longer, for some better qualified system ... with a Ph.D.?” (1987, p. 117).

This line of thought was discussed by Eugene Wigner, “This way out of the difficulty amounts to the postulate that the equations of motion of quantum mechanics cease to be linear, in fact that they are grossly non-linear if conscious beings enter the picture” (1967, p. 183). Wigner suggests that the quantum measurement problem indicates “the effect of consciousness on physical phenomena,” a possibility of almost incomprehensible implications (not the least of which: How could conscious beings evolve if there were no collapses, since the universe would surely be in a superposition of states with and without conscious beings!). In any case, Wigner’s speculations never amounted to a physical theory, nor could they unless a physical characterization of a conscious system was forthcoming.

So if one adopts a collapse theory, and if the collapses are tied to measurements or observations, then one is left with the problem of giving a physical characterization of an observation or a measurement. Such physicists as Einstein and Bell were incredulous of the notion that conscious systems play such a central role in the physics of the universe.

SPONTANEOUS COLLAPSE THEORIES. Nonetheless, precise theories of collapse do exist. The key to resolving the foregoing puzzle is to notice that although collapses must be of the right form to make the physical interactions called “observations” and “measurements” have determinate outcomes, there is no reason that the collapse dynamics itself need mention observation or measurement. The collapse dynamics merely must be of such a kind as to give outcomes in the right situations.

The most widely discussed theory of wave function collapse was developed by Gian Carlo Ghirardi, Alberto Rimini, and Tulio Weber (1986) and is called the spontaneous localization theory or, more commonly, the GRW theory. The theory postulates an account of wave function collapse that makes no mention of observation, measurement, consciousness, or anything of the sort. Rather, it supplies a universal rule for both how and when the collapse occurs. The “how” of the collapse involves localization in space; when the collapse occurs, one takes a single particle and multiplies its wave function, expressed as a function of space, by a narrow Gaussian (bell curve). This has the effect of localizing the particle near the center of the Gaussian, in the sense that most of the wave function will be near the center. If the wave function before the collapse is widely spread out over space, after the collapse it is much more heavily weighted to a particular region. The likelihood that a collapse will occur centered at a particular location depends on the square amplitude of the precollapse wave function for that location. The collapses, unlike Schrödinger evolution, are fundamentally nondeterministic, chancy events.

The GRW collapse does not perfectly locate the wave function at a point. It could not do so for straightforward physical reasons: The localization process will violate the conservation of energy, and the more narrowly the post-collapse wave function is confined, the more new energy is pumped into the system. If there were perfect localizations, the energy increase would be infinite—and immediately evident. (It follows from these same observations that even in the “standard” theory there are never collapses to perfectly precise positions—even after a so-called position measurement.)

Therefore, the GRW theory faces a decision: Exactly how localized should the localized wave function be? This corresponds to choosing a width for the Gaussian: The narrower the width, the more energy that is added to the system on collapse. The choice for this width is bounded in one direction by observation—the energy increase for the universe must be below observed bounds, and particular processes, such as spontaneous ionization, should be

rare—and in the other direction by the demand that the localization solve the measurement problem. As it happens, Ghirardi, Rimini, and Weber chose a value of about  $10^{-5}$  centimeters for the width of the Gaussian. This is a new constant of nature.

Beside the “how” of the collapse, the GRW theory must specify the “when.” It was here that we saw issues such as consciousness getting into the discussion: If collapses occur only when measurements or observations occur, then we must know when measurements or observations occur. The GRW theory slices through this problematic neatly; it simply postulates that the collapses take place at random, with a fixed probability per unit time. This introduces another new fundamental constant: the average time between collapses per particle. The value of that constant is also limited in two directions; on the one hand, we know from interference experiments that isolated individual particles almost never suffer collapses on the time scale of laboratory operations. On the other hand, the collapses must be frequent enough to resolve the measurement problem. The GRW theory employs a value of  $10^{15}$  seconds, or about 100 million years, for this constant.

Clearly, the constant has been chosen large enough to solve one problem: Individual isolated particles will almost never suffer collapses in the laboratory. It is less clear, though, how it solves the measurement problem.

The key here is to note that actual experiments record their outcomes in the correlated positions of many, many particles. In our spin experiment we said that our spin measuring device must have two distinct indicator states: |“up” $\rangle$  and |“down” $\rangle$ . To be a useful measuring device, these indicator states must be macroscopically distinguishable. This is achieved with macroscopic objects—pointers, drops of ink, and so on—to indicate the outcome. And a macroscopic object will have on the order of  $10^{23}$  particles.

So suppose the outcome |“up” $\rangle$  corresponds to a pointer pointing to the right and the outcome |“down” $\rangle$  corresponds to the pointer pointing to the left. If there are no collapses, the device will end up with the wave function  $1/\sqrt{2}|x\text{-up}\rangle_s|“up”\rangle_A + 1/\sqrt{2}|x\text{-down}\rangle_s|“down”\rangle_A$ . Now although it is unlikely that any particular particle in the pointer will suffer a collapse on the time scale of the experiment, because there are so many particles in the pointer, it is overwhelmingly likely that some particle or other in the pointer will suffer a collapse quickly: within about  $10^{-8}$  seconds. And (this is the key), since in the state  $1/\sqrt{2}|x\text{-up}\rangle_s|“up”\rangle_A + 1/\sqrt{2}|x\text{-down}\rangle_s|“down”\rangle_A$  all the particle positions are correlated with one another, if the

collapse localizes a single particle in the pointer, it localizes all of them. So, if having the wave functions of all the particles in the pointer highly concentrated on the right (or on the left) suffices to solve the measurement problem, the problem will be solved before  $10^{-4}$  seconds has elapsed.

The original GRW theory has been subject to much discussion. In a technical direction there have been similar theories, by Ghirardi and Rimini and by Philip Pearle, that make the collapses to be continuous rather than discrete. More fundamentally, there have been two foundational questions: First, does the only approximate nature of the “localization” vitiate its usefulness in solving the measurement problem, and second, does the theory require a physical ontology distinct from the wave function? Several suggestions for such an additional ontology have been put forward, including a mass density in space-time, and discrete events (“flashes”) in space-time.

The addition of such extra ontology, beyond the wave function, reminds us of the second horn of Bell’s dilemma: Either the wave function as given by the Schrödinger equation is not right or it is not everything. The versions of the GRW theory that admit a mass density or the flashes postulate that the wave function is not everything, do so in such a way that the exact state of the extra ontology can be recovered from the wave function. The more radical proposal is that there is extra ontology, and its state cannot be read off the wave function. These are the so-called hidden variables theories.

### ADDITIONAL VARIABLES THEORIES

According to an additional variables theory, the complete quantum state of the system after a measurement is indeed  $1/\sqrt{2}|x\text{-up}\rangle_S|“\text{up}”\rangle_A + 1/\sqrt{2}|x\text{-down}\rangle_S|“\text{down}”\rangle_A$ . The outcome of the measurement cannot be read off of that state because the outcome is realized in the state of the additional variables, not in the wave function. It immediately follows that for any such theory, the additional ontology, the additional variables, had best not be “hidden”: since the actual outcome is manifest, the additional variables had best be manifest. Indeed, on this approach the role of the wave function in the theory is to determine the evolution of the additional variables. The wave function, since it is made manifest only through this influence, is really the more “hidden” part of the ontology.

The best known and most intensively developed additional variables theory goes back to Louis de Broglie, but is most intimately associated with David Bohm. In its nonrelativistic particle version, Bohmian mechanics,

physical objects are constituted of always-located point particles, just as was conceived in classical mechanics. At any given time, the physical state of a system comprises both the exact positions of the particles and a wave function. The wave function never collapses: it always obeys a linear dynamical equation like the Schrödinger equation. Nonetheless, at the end of the experiment the particles in the pointer will end up either all on the right or all on the left, thus solving the measurement problem. This is a consequence of the dynamics of the particles as determined by the wave function.

It happens that the particle dynamics in Bohmian mechanics is completely deterministic, although that is not fundamentally important to the theory and indeterministic versions of Bohm’s approach have been developed. The dynamical equation used in Bohmian mechanics is much more importantly the simplest equation that one can write down if one assumes that the particle trajectories are to be determined by the wave function and that various symmetries are to be respected. If one starts with idea that there are particles and that quantum theory should be a theory of the motion of those particles that reproduces the predictions of the standard mathematical recipe, Bohmian mechanics is the most direct outcome.

Since Bohmian mechanics is a deterministic theory, the outcome of any experiment is fixed by the initial state of the system. The probabilities derived from the standard mathematical recipe must therefore be interpreted purely epistemically: they reflect our lack of knowledge of the initial state. This lack of knowledge turns out to have a physical explanation in Bohmian mechanics: Once one models any interaction designed to acquire information about a system as a physical interaction between a system and an observer, it can be shown to follow that initial uncertainty about the state of the target system cannot be reduced below a certain bound, given by the Heisenberg uncertainty relations.

This illustrates the degree to which the ontological “morals” of quantum theory are held hostage to interpretations. In the standard interpretation, when the wave function of a particle is spread out, there is no further fact about exactly where the particle is. (Because of this, position measurements in the standard theory are not really measurements, i.e., they do not reveal preexisting facts about positions.) In Bohm’s interpretation, when the wave function is spread out, there is a fact about exactly where the particle is, but it follows from physical analysis that one cannot find out more exactly where it is without thereby altering the wave function (more properly, with-



out altering the effective wave function that we use to make predictions). Similarly, in the standard interpretation, when we do a position measurement on a spread out particle, there is an indeterministic collapse that localizes the particle—it gives it an approximate location. According to Bohm’s theory the same interaction really is a measurement: It reveals the location that the particle already had. So it is a fool’s errand to ask after “the ontological implications of quantum theory”: the account of the physical world one gets depends critically on the interpretation of the formalism.

Bohm’s approach has been adapted to other choices for the additional variables. In particular, interpretations of field theory have been pursued in two different ways: with field variables that evolve indeterministically, and with the addition to Bohmian mechanics the possibility of creating and annihilating particles in an indeterministic way. Each of these provides the wherewithal to treat standard field theory.

There have been extensive examinations of other ways to add additional variables to a noncollapse interpretation, largely under the rubric of *modal interpretations*. Both rules for specifying what the additional variables are and rules for the dynamics of the new variables have been investigated.

### A THIRD WAY?

There are also some rather radical attempts to reject each of Bell’s two options and to maintain both that the wave function, as given by the Schrödinger equation, is right and that it is everything—that is, it is descriptively complete. Since a wave function such as  $1/\sqrt{2}|x\text{-up}\rangle_s|“\text{up}”\rangle_A + 1/\sqrt{2}|x\text{-down}\rangle_s|“\text{down}”\rangle_A$  does not indicate that one outcome rather than the other occurred, this requires maintaining that it is not the case that one outcome rather than the other occurred.

This denial can come in two flavors. One is to maintain that neither outcome occurred, or even seemed to occur, and one is only somehow under the illusion that one did. David Z. Albert (1992) investigated this option under the rubric the *bare theory*. Ultimately, the bare theory is insupportable, since any coherent account must at least allow that the quantum mechanical predictions appear to be correct.

The more famous attempt in this direction contends that, in some sense, both outcomes occur, albeit in different “worlds.” Evidently, the wave function  $1/\sqrt{2}|x\text{-up}\rangle_s|“\text{up}”\rangle_A + 1/\sqrt{2}|x\text{-down}\rangle_s|“\text{down}”\rangle_A$  can be written as the mathematical sum of two pieces, one of

which corresponds to a situation with the apparatus indicating “up” and the other to a situation with the apparatus indicating “down.” The *many worlds theory* attempts to interpret this as a single physical state, which somehow contains or supports two separate “worlds,” one with each outcome.

The many worlds interpretation confronts several technical and interpretive hurdles. The first technical hurdle arises because any wave function can be written as the sum of other wave functions in an infinitude of ways. For example, consider the apparatus state  $1/\sqrt{2}|“\text{up}”\rangle_A + 1/\sqrt{2}|“\text{down}”\rangle_A$ . Intuitively, this state does not represent the apparatus as having fired one way or another. This state can be called  $|D_1\rangle_A$ . Similarly,  $|D_2\rangle_A$  can represent the state  $1/\sqrt{2}|“\text{up}”\rangle_A - 1/\sqrt{2}|“\text{down}”\rangle_A$ , which also does not correspond to an apparatus with a definite outcome. The state  $1/\sqrt{2}|x\text{-up}\rangle_s|“\text{up}”\rangle_A + 1/\sqrt{2}|x\text{-down}\rangle_s|“\text{down}”\rangle_A$ , which seems to consist in two “worlds,” one with each outcome, can be written just as well as  $1/\sqrt{2}|y\text{-up}\rangle_s|D_1\rangle_A + 1/\sqrt{2}|y\text{-down}\rangle_s|D_2\rangle_A$ . Written in this way, the state seems to comprise two worlds: one in which the electron has  $y$ -spin up and the apparatus is not in a definite indicator state, the other in which the electron has  $y$ -spin down, and the apparatus is in a distinct physical state that is equally not a definite indicator state. If these are the “two worlds,” then the measurement problem has not been solved, it has been merely traded as a single world without a definite outcome for a pair of worlds neither of which has a definite outcome.

So the many worlds theory would first have to maintain that there is a preferred way to decompose the global wave function into “worlds.” This is known as the *preferred basis problem*.

A more fundamental difficulty arises when one tries to understand the status of the probabilities in the many worlds theory. In a collapse theory the probabilities are probabilities for collapses to occur one way rather than another, and there is a physical fact about how the collapses occur, and therefore about frequencies of outcomes. In an additional variables theory the probabilities are about which values the additional variables take, and there is a physical fact about the values they take and therefore about frequencies of outcomes. But in the many worlds theory, whenever one does an experiment like the spin measurement described earlier, the world splits: There is no frequency with which one outcome occurs as opposed to the other. And more critically, that the world “splits” has nothing to do with the amplitude assigned to the two daughter worlds.

Suppose, for example, that instead of feeding a  $y$ -spin up electron into our  $x$ -spin measuring device, we feed in an electron whose state is  $1/2|x\text{-up}\rangle_s + \sqrt{3}/2|x\text{-down}\rangle_s$ . By linearity, at the end of the experiment, the state of the system plus apparatus is  $1/2|x\text{-up}\rangle_s|“up”\rangle_A + \sqrt{3}/2|x\text{-down}\rangle_s|“down”\rangle_A$ . Even if we have solved the preferred basis problem and can assert that there are now two worlds, one with each outcome, notice that we are evidently in exactly the same situation as in the original experiment: Whenever we do the experiment, the universe “splits.” But the quantum formalism counsels us to have different expectations in the two cases: in the first case, we should expect to get an “up” outcome 50 percent of the time, in the second case only 25 percent of the time. It is unclear, in the many worlds theory, what the expectations are for, and why they should be different.

Another interpretation of the quantum formalism that has been considered is the *many minds theory* of Barry Loewer and Albert. Despite the name, the many minds theory is not allied in spirit with the many worlds theory: It is rather an additional variables theory in which the additional variables are purely mental subjective states. This is somewhat akin to Wigner’s appeal to consciousness to solve the measurement problem, but where Wigner’s minds affect the development of the wave function, the minds in this theory (as is typical for additional variables theories) do not. The physical measurement apparatus in the problematic case does not end up in a definite indicator state, but a mind is so constituted that it will, in this situation, have the subjective experience of seeing a particular indicator state. Which mental state the mind evolves into is indeterministic. The preferred basis problem is addressed by stipulating that there is an objectively preferred basis of physical states that are associated with distinct mental states.

The difference between the many worlds and the many minds approaches is made most vivid by noting that the latter theory does not need more than one mind to solve the measurement problem, where the problem is now understood as explaining the determinate nature of our experience. A multiplicity of minds are added to Loewer and Albert’s theory only to recover a weak form of mind-body supervenience: Although the experiential state of an individual mind does not supervene on the physical state of the body with which it is associated, if one associates every body with an infinitude of minds, the distribution of their mental states can supervene on the physical state of the body.

A final attempt to address the problems of quantum mechanics deserves brief mention. Some maintain that

the reason quantum mechanics is so confusing is not because the mathematical apparatus requires emendation (e.g., by explicitly adding a collapse or additional variables) or an interpretation (i.e., an account of exactly which mathematical objects represent physical facts), but because we reason about the quantum world in the wrong way. Classical logic, it is said, is what is leading us astray. We merely need to replace our patterns of inference with quantum logic.

There is a perfectly good mathematical subject that sometimes goes by the name *quantum logic*, which is the study, for example, of relations between subspaces of Hilbert space. These studies, like all mathematics, employ classical logic. There is, however, no sense in which these studies, by themselves, afford a solution to the measurement problem or explain how it is that experiments like those described earlier have unique, determinate outcomes.

## THE WAVE FUNCTION, ENTANGLEMENT, EPR, AND NON-LOCALITY

For the purposes of this discussion, the wave function has been treated as if it were something like the electromagnetic field: a field defined on space. Although this is not too misleading when discussing a single particle, it is entirely inadequate when considering collections of particles. The wave function for  $N$  particles is a function not on physical space, but on the  $3N$ -dimensional configuration space, each point of which specifies the exact location of all the  $N$  particles. This allows for the existence of entangled wave functions, in which the physical characteristics of even widely separated particles cannot be specified independently of one another.

Consider  $R$  and  $L$ , a pair of widely separated particles. Among the wave functions available for this pair is one that ascribes  $x$ -spin up to  $R$  and  $x$ -spin down to  $L$ , which is written as  $|x\text{-up}\rangle_R|x\text{-down}\rangle_L$ , and one that attributes  $x$ -spin down to  $R$  and  $x$ -spin up to  $L$ :  $|x\text{-down}\rangle_R|x\text{-up}\rangle_L$ . These are called *product states*, and all predictions from these states about how  $R$  will respond to a measurement are independent of what happens to  $L$ , and vice versa.

But besides these product states, there are entangled states like the *singlet state*:  $1/\sqrt{2}|x\text{-up}\rangle_R|x\text{-down}\rangle_L - 1/\sqrt{2}|x\text{-down}\rangle_R|x\text{-up}\rangle_L$ . In this state the  $x$ -spins of the two particles are said to be anticorrelated since a measurement of their  $x$ -spins will yield either up for  $R$  and down for  $L$  or down for  $R$  and up for  $L$  (with a 50 percent chance for each outcome). Even so, if the wave function is complete, then neither particle in the singlet state has a

determinate  $x$ -spin: the state is evidently symmetrical between spin up and spin down for each particle considered individually.

How can the  $x$ -spins of the particles be anticorrelated if neither particle has an  $x$ -spin? The standard answer must appeal to dispositions: although in the singlet state neither particle is disposed to display a particular  $x$ -spin on measurement, the pair is jointly disposed to display opposite  $x$ -spins if both are measured. Put another way, on the standard interpretation, before either particle is measured neither has a determinate  $x$ -spin, but after one of them is measured, and, say, displays  $x$ -spin up, the other acquires a surefire disposition to display  $x$ -spin down. And this change occurs simultaneously, even if the particles happen to be millions of miles apart.

Einstein found this to be a fundamentally objectionable feature of the standard interpretation of the wave function. In a paper coauthored with Boris Podolsky and Nathan Rosen (EPR 1935), Einstein pointed out this mysterious, instantaneous “spooky action-at-a-distance” built into the standard approach to quantum theory. It is uncontroversial that an  $x$ -spin measurement carried out on L with, say, an “up” outcome” will result in a change of the wave function assigned to R: It will now be assigned the state  $|x\text{-down}\rangle_R$ . If the wave function is complete, then this must reflect a physical change in the state of R because of the measurement carried out on L, even though there is no physical process that connects the two particles. What EPR pointed out (using particle positions rather than spin, but to the same effect) was that the correlations could easily be explained without postulating any such action-at-a-distance. The natural suggestion is that when we assign a particular pair of particles the state  $1/\sqrt{2}|x\text{-up}\rangle_R|x\text{-down}\rangle_L - 1/\sqrt{2}|x\text{-down}\rangle_R|x\text{-up}\rangle_L$ , it is a consequence of our ignorance of the real physical state of the pair: The pair is either in the product state  $|x\text{-up}\rangle_R|x\text{-down}\rangle_L$  or in the product state  $|x\text{-down}\rangle_R|x\text{-up}\rangle_L$ , with a 50 percent chance of each. This simple expedient will predict the same perfect anticorrelations without any need to invoke a real physical change of one particle consequent to the measurement of the other.

So matters stood until 1964, when Bell published his famous theorem. Bell showed that Einstein’s approach could not possibly recover the full range of quantum mechanical predictions. That is, no theory can make the same predictions as quantum mechanics if it postulates (1) that distant particles, such as R and L, have each their own physical state definable independently of the other and (2) measurements made on each of the particles have

no physical affect on the other. Entanglement of states turns out to be an essential feature—arguably the central feature—of quantum mechanics. And entanglement between widely separated particles implies non-locality: The physics of either particle cannot be specified without reference to the state and career of the other.

The spooky action-at-a-distance that Einstein noted is not just an artifact of an interpretation of the quantum formalism; it is an inherent feature of physical phenomena that can be verified in the laboratory. A fundamental problem is that the physical connection between the particles is not just spooky (unmediated by a continuous space-time process), it is superluminal. It remains unclear to this day how to reconcile this with the theory of relativity.

**See also** Bohm, David; Bohmian Mechanics; Many Worlds/Many Minds Interpretation of Quantum Mechanics; Modal Interpretation of Quantum Mechanics; Non-locality; Philosophy of Physics; Quantum Logic and Probability.

### Bibliography

- Albert, David Z. *Quantum Mechanics and Experience*. Cambridge, MA: Harvard University Press, 1992.
- Bell, John S. *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy*. Cambridge, U.K.: Cambridge University Press, 1987.
- Dürr, Detlef, Sheldon Goldstein, and Nino Zanghi. “Quantum Equilibrium and the Origin of Absolute Uncertainty.” *Journal of Statistical Physics* 67 (1992): 843–907.
- Ghirardi, GianCarlo, Alberto Rimini, and Tullio Weber. “Unified Dynamics for Microscopic and Macroscopic Systems.” *Physical Review* 34 (2) (1986): 470–491.
- Maudlin, Tim. *Quantum Non-locality and Relativity: Metaphysical Intimations of Modern Physics*. Malden, MA: Blackwell, 2002.
- Von Neumann, John. *Mathematical Foundations of Quantum Mechanics*. Translated by Robert T. Beyer. Princeton, NJ: Princeton University Press, 1955.
- Wheeler, John Archibald, and Wojciech Hubert Zurek, eds. *Quantum Theory and Measurement*. Princeton, NJ: Princeton University Press, 1983.
- Wigner, Eugene. *Symmetries and Reflections*. Westport, CT: Greenwood Press, 1967.

**Tim Maudlin (2005)**

## QUESTIONS

See Appendix, Vol. 10

## QUINE, WILLARD VAN ORMAN

(1908–2000)

Willard Van Orman Quine, an Edgar Pierce professor of philosophy at Harvard, was born in Akron, Ohio. In 1930 he was graduated from Oberlin, where he majored in mathematics, and he wrote a doctoral dissertation in logic under Alfred North Whitehead at Harvard. He visited Vienna, studied mathematical logic at Warsaw, and at Prague met Rudolf Carnap, whose work was to inspire and influence him.

Some of Quine's publications are in philosophy, some in symbolic logic, and others are concerned with the logical regimentation of ordinary language. It is his philosophy and related aspects of his advocated regimentation of language that concern us here, his contributions to logic being dealt with elsewhere.

### ANALYTIC-SYNTHETIC DISTINCTION

Some philosophers have attempted to distinguish between such statements as "A river flows through Brisbane," which, they contend, are true as a matter of fact, and statements like "No bachelor is married," the truth of which is said to be independent of matters of fact. The former have been described as synthetic, the latter as analytic. Quine maintained, first, that the analytic-synthetic distinction has never satisfactorily been made and, second, that there is no good reason for believing that it can be made.

**LOGICAL TRUTH.** Given a list of logical particles and the notion of truth, with which Quine was comparatively satisfied, we may, he contends, derive the notion of logical truth. "All birds are birds" is logically true because it is both true and such that if we leave its logical parts alone and replace "birds" with some other word, then if we get a statement at all, we get a true one—for example, "All snakes are snakes." But even though this analytic statement is logically true, there are analytic statements like "No bachelor is married" that are not, and thus analyticity remains to be explained. If we replace "bachelor" with the synonymous "unmarried man," we have a logical truth, and it would thus appear that an analytic statement either is a logical truth or is reducible to one by interchange of synonyms.

**SYNONYMY.** However, according to Quine, an account of analyticity that depends on the notion of synonymy is unsatisfactory. Suppose that all and only Guards officers

are very tall soldiers with long hair. Since "Guards officers" and "very tall soldiers with long hair" are coextensive expressions, there are statements whose truth or falsity cannot be affected by interchanging these expressions. But because they are not synonymous expressions, there are also statements like "Necessarily, all and only Guards officers are Guards officers" that can be so affected. In contrast, the truth of the statement "Necessarily, all and only bachelors are bachelors" cannot be affected by interchanging "bachelors" and "unmarried men" because these expressions are synonymous. But to make the last statement is to say that "All and only bachelors are bachelors" is analytic. Thus, we give an account of synonymy in terms of the effects of interchanging expressions in certain contexts. But because these contexts cannot be specified without reference to analyticity or some equivalent notion, we cannot, without circularity, use the notion of synonymy in giving an account of analyticity. Similar difficulties frustrate the derivation of self-contradictoriness from logical falsity.

Quine also discusses the possibility of giving an account of the analyticity of statements in artificial languages, but here, as in natural languages, the difficulty is, he contended, that each of the key notions in the theory of meaning is definable only in terms of the others.

Anyone who produced an account of these notions acceptable to Quine would thereby refute him, but what sort of account this would be remains to be seen. In the meantime the strongest argument against him is *ad hominem*. "All the illuminated manuscripts are illuminated" is logically true only if "illuminated" has the same meaning in each of its occurrences. Thus, the notion of logical truth, which Quine accepts, is dependent upon the notion of synonymy, which he rejects.

**RADICAL TRANSLATIONS.** Quine's theory of meaning was further developed in his discussion of the difficulties that would arise if we were to attempt to translate the language of a hitherto isolated tribe. Radical translation, as he calls it, would have to begin not with words but with those sentences that have a comparatively direct relation to stimulus conditions. The stimulus meaning of a sentence for a person is defined in terms of the class that has as its members the kinds of stimulation that would prompt the person's assent to the sentence. Intrasubjective stimulus synonymy is sameness of stimulus meaning for one speaker, and two sentences are socially stimulus-synonymous if they are intrasubjectively stimulus-synonymous for nearly everyone who speaks the language. A sentence is stimulus-analytic for a person if he would

assent to it, if to anything, after every stimulation, and a socially stimulus-analytic sentence is stimulus-analytic for nearly every speaker of the language.

In order to see that these are not our intuitive notions of synonymy and analyticity, we need to distinguish occasion sentences and standing sentences. If every minute or so we are asked to assent to “John has hiccups,” we cannot do so without having another look at John on each occasion. In contrast, having once assented to the standing sentence “Salt is soluble in water,” we may assent again without observing salt or anything else again. Applied to occasion sentences, intrasubjective stimulus synonymy approximates sameness of meaning; standing sentences, however, are related to experience indirectly, and the kinds of stimulus that would prompt assent to a standing sentence vary from speaker to speaker. Thus, the stimulus meaning of a standing sentence falls short of our intuitive notion of meaning; stimulus synonymy is correspondingly inadequate, and some socially stimulus-analytic sentences would normally be described not as analytic but as conveying information common to the whole community.

Quine demands of those who talk of analyticity and synonymy that they give of their concepts the sort of account in terms of dispositions to verbal behavior that he has given of his.

By observing and testing native speech behavior dispositions, the linguist can come to translate some occasion sentences and to recognize stimulus analyticity and synonymy. But in order to complete the radical translation of a language, he must frame analytical hypotheses. This consists of segmenting what he hears into native words and hypothetically equating these to English expressions. Quine contends that there will be many sets of analytical hypotheses that fit all native dispositions to speech behavior and yet lead to incompatible translations of countless sentences in their language. Suppose that, observing the circumstances in which a native utters “Gavagai,” we translate this sentence as “Rabbit!” Whether the word *gavagai* is to be taken to apply to rabbits, temporal stages of rabbits, or something even stranger to us can be settled only when we can ask questions like “Is this the same rabbit as that?” This cannot be done until we have translated the parts of speech that make up the native system of reference, and since this is part of what we do when we adopt a set of analytical hypotheses, there is more than one way of doing it. For example, the sentence translated as “Is this (the same) (rabbit) as that?” might, on another set of empirically sat-

isfactory hypotheses, be translated as “Is this (a rabbit stage) (of the same series) as that?”

In this way Quine arrives at the principle of the indeterminacy of translation, which says that it is possible to compile incompatible manuals for translating one language into another, all of which fit all observable speech dispositions, and that there is no sense in asking which is the right manual. It is only in exceptional cases that we can talk of the meaning of a single sentence, and when our statements about the world conflict with experience, they do so not individually but as a system. Thus, we have what might be called the Quine-Duhem conventionalist thesis that any statement can be held to be true no matter what is observed, provided that adjustments are made elsewhere in the system; it is from this thesis that Quine infers that it is impossible to make the analytic-synthetic distinction.

Quine believed that his discussion of radical translation reveals the possibility of differences between the conceptual schemes of people that are not empirically conditioned. In the case of two compatriot linguists working independently on the radical translation of a language, one linguist might conclude that he and the native see the world in the same way, as consisting of tables, chairs, ducks, and rabbits, while the other finds that the native speaks of rabbit stages, not of rabbits, and concludes that the native’s outlook is different from his own. Now, in order to determine what the native’s outlook really is, it is necessary to discover which is the correct way of translating the native’s language. But according to the principle of the indeterminacy of translation, it does not make sense even to ask this, and consequently it cannot make sense to ask what the native’s outlook is. It can be shown that the native is in no better position than the linguist here, and it then becomes hard to see the sense of talking about an outlook when there is no conceivable way of discovering what this outlook is. Quine’s position here is not clear. He admitted that these differences of outlook are in principle undetectable and grants that such cultural contrasts are threatened with meaninglessness, but he continued to speak of them.

As radical translation is not known ever to have been undertaken, the absence of incompatible manuals of translation does not count against the principle of indeterminacy. Nevertheless, it might well be contended that until there are more conclusive arguments for it, the principle is to be taken as the incredible consequence of unsound premises. Quine, in discussing meaning, did concentrate on the statement-making function of language, and it has, in fact, been argued that by neglecting

the countless other uses of language, he arrived at a concept of synonymy the inadequacy of which is revealed by the fact that it makes translation indeterminate.

## ONTOLOGY

Philosophers have disagreed as to what there is; some have held, for example, that there are only material things, and others have denied this. Quine called such theories “ontic theories” and maintained that they are a part of the sciences distinguished only by extreme generality. Given that there are physical objects, it is the natural scientist who discovers whether there are wombats; and given classes, it is the mathematician who finds out whether there are even prime numbers. Whether there are physical objects and classes, however, is the concern of the philosopher. The integration of established theories, which is one of the aims of scientific work, may lead to any one of many equally satisfactory accounts of the world, each with its ontic theory, and there is no sense in asking which of these accounts is the true one. Thus, Quine took a conventionalist view even of the theses of ontologists.

Today it is commonly maintained that since there is no way of settling an ontic dispute, ontologists have unwittingly concerned themselves with pseudo questions. Quine, in proposing a method of determining the ontic import of a theory, attempted to make such questions decidable and thus real. His method was, in outline, as follows: “ $(\exists x)(x \text{ is a cat})$ ” may be read as “There is an  $x$  such that  $x$  is a cat” or as “There is something such that it is a cat.” According to Quine, anyone who makes this statement is thereby committed to the existence of cats. The statement consists of the existential quantifier “ $(\exists x)$ ,” the predicate “— is a cat,” and an “ $x$ ” that works like a pronoun and is needed in any but the simplest cases to show under which quantifier a predicate comes. If we add to this equipment such truth-functional words as “*and*” and “*not*,” we can make statements like “ $(\exists x)(x \text{ is a book, and } x \text{ is boring})$ , and  $(\exists x)(x \text{ is a book, and } x \text{ is not boring})$ .” This is a paraphrase of “Some but not all books are boring,” which, it is alleged, reveals the ontic import of this statement. Bertrand Russell, Quine, and others have suggested similarly revealing paraphrases of general hypotheticals, of statements containing proper names, and of statements containing such descriptive phrases as “the prime number between 5 and 11.” Quine contended that in adopting any theory, we commit ourselves to the existence of certain entities and that by translating the theory into a language in which the only formal devices

are predication, quantification, and truth-functional composition, we make these commitments explicit.

**ONTIC COMMITMENTS.** The commitments revealed in the above manner are incurred when certain words are used in certain ways. We are, according to Quine, committed to the existence of physical objects because of the ways in which physical object terms function in our language. In contrast, we are not committed to such objects as “sakes,” because even though we do some things for the sake of others, “sake” functions in only a few of the ways in which a term does. When constructing theories, we are, within limits, free to decide what expressions will function as terms, and by such decisions we might commit ourselves to the existence of atoms, for example, but not to that of meters. We accept the reality of physical objects more readily than we do that of atoms because typical sentences about physical objects are more closely associated with sensory stimulation than are typical sentences about atoms. By this criterion sense data are even more acceptable than physical objects, but this is counteracted by the fact that sense data are a less satisfactory basis for an account of the world. On the grounds of utility for theory, classes are to be preferred to attributes and sentences to propositions.

Many would maintain that it is only when Quine is discussing the considerations that influence ontic decisions that he tackles philosophical problems, and that he does this in a way he himself admitted to be sketchy. He does this sketchily because it has been done in detail by others to whom he refers, and believing that ontologists must take account of scientific theories, he is especially interested in working out how this is to be done. Perhaps the major philosophical problem raised by Quine’s proposed criterion of ontic commitment is that of the nature of this commitment: I may know what it is like for a nation to be, or not to be, committed to an isolationist foreign policy, but what is it like to be, or not to be, committed to the existence of physical objects?

**REGIMENTATION OF ORDINARY LANGUAGE.** The regimentation of language serves purposes other than that of revealing ontic commitments. The logic of ordinary language is difficult to formulate, and consequently it is more economical to theorize in a language that is ordinary except in its logical parts, which are designed to facilitate deduction. And if there are fewer kinds of construction and less obscurity in a regimented language, then in moving into it we simplify and clarify our conceptual scheme.

Because of misgivings about synonymy Quine cannot maintain that for an ordinary-language sentence to be replaced by a regimented one, the two must be synonymous. Indeed, we may be making the replacement just because one sentence is ambiguous and the other is not. Paraphrase into a regimented language consists, he maintains, of replacements that, in certain contexts, forward certain programs. Against this it has been argued that for any two sentences there will be a program that is forwarded by replacing one with the other, and consequently Quine's notion of paraphrase is vacuous unless contexts and programs can be specified. If this can be done, however, the notion of sentence synonymy can be derived. This notion is no less satisfactory, and no more difficult to make adequate sense of, than the notion of paraphrase, without which Quine cannot talk of putting theories into a regimented language.

The bulk of Quine's philosophical work was published after 1947. By 1960 he had combined into a coherent position theses some of which were first put forward ten years earlier. Between 1947 and 1960 certain changes in his views occurred. From declaring, in 1947, that he did not believe in abstract entities, he had come not only to accept such entities but also to claim that he had always done so; from counting phenomenalism, in 1948, as a conceptual scheme suitable for certain purposes, he came to reject it; and from maintaining, in 1951, that in the face of recalcitrant experience we could change our logical laws, he had apparently come to hold that there is nothing that would count as changing our logical laws.

Quine's status as a philosopher never depended upon the number of people who agreed with him. On the contrary, the sign of his achievement is the valuable discussion he provoked by his persistent and penetrating attacks on analyticity and related notions and by his unfashionable conviction that philosophers want to discover what reality is like.

**See also** Analytic and Synthetic Statements; Analyticity; Artificial and Natural Languages; Carnap, Rudolf; Logic, History of; Ontology; Philosophy of Language; Synonymy; Underdetermination Thesis, Duhem-Quine Thesis; Whitehead, Alfred North.

### Bibliography

The topics of the two main sections above were first treated in detail in "Two Dogmas of Empiricism" (1951) and "On What There Is" (1948), reprinted in Quine's *From a Logical Point of View* (Cambridge, MA: Harvard University Press, 1953). A definitive statement of Quine's position, including an exposition of the indeterminacy thesis, is in his *Word and*

*Object* (New York: MIT Press, 1960), in which a list of Quine's many other philosophical publications will be found. His views on analyticity and logical truth are briefly stated in his "Mr. Strawson on Logical Theory," in *Mind* 62 (1953): 433–451. Important earlier articles are his "Truth by Convention," in *Philosophical Essays for Alfred North Whitehead* (New York: Longmans, Green, 1936), reprinted in Herbert Feigl and Wilfrid Sellars, eds., *Readings in Philosophical Analysis* (New York: Appleton-Century-Crofts, 1949), and "Steps toward a Constructive Nominalism," in *Journal of Symbolic Logic* 12 (1947): 105–122, which he wrote with Nelson Goodman.

Among the many discussions of Quine's views are the following, which are the sources of critical points made in this article. H. P. Grice and P. F. Strawson, "In Defence of a Dogma," in *Philosophical Review* 65 (1956): 141–158, defend the analytic-synthetic distinction. P. F. Strawson, "Propositions, Concepts and Logical Truths," in *Philosophical Quarterly* 7 (1957): 15–25, is an attack on Quine's notion of logical truth. Comments on indeterminacy are made in L. J. Cohen, *The Diversity of Meaning* (London: Methuen, 1962), pp. 67–74. Quine's views on ontology are criticized by G. J. Warnock in "Metaphysics in Logic," in *Essays in Conceptual Analysis*, edited by A. G. N. Flew (London: Macmillan, 1956), pp. 75–93, and from a different standpoint in Rudolf Carnap, *Meaning and Necessity* (Chicago: University of Chicago Press, 1947). The relation between conventionalism and the analytic-synthetic distinction is discussed in G. H. Herbert, "The Analytic and the Synthetic," in *Philosophy of Science* 26 (1959): 104–113. Some points made above are also developed in C. F. Presley, "Quine's *Word and Object*," in *Australasian Journal of Philosophy* 39 (1961): 175–190.

C. F. Presley (1967)

## QUINE, WILLARD VAN ORMAN [ADDENDUM]

Willard Van Orman Quine, the Edgar Pierce Professor of Philosophy Emeritus, at Harvard, author of twenty-one books and scores of journal articles and reviews, made many significant contributions to metaphysics, epistemology, philosophy of language, philosophy of science, philosophy of mind, logic, philosophy of logic, and set theory, and ethics (and ethical theory). These contributions are of a stature that firmly places Quine among the titans of twentieth-century Anglo American philosophy.

In most of his publications following *Word and Object* (1960), Quine sought to sum up, clarify, and expand on various themes found in that book. Quine can occasionally be seen changing his mind regarding some detail of his prior thought, but by and large he remains remarkably consistent.

## NATURALISM

The keystone of Quine's systematic philosophy is naturalism. Roughly, naturalism is the view that there is no suprascientific justification for science *and* that it is up to science to determine both what there is (ontology) and how we know what there is (epistemology). Moreover, Quine maintains that the best current science tentatively and fallibly plumps for a physicalist ontology and an empiricist epistemology.

**ONTOLOGY: PHYSICALISM.** Since he maintains that what a (formalized) theory says there is is determined by the range of values of the bound variables of that theory, and since the bound variables of the best current scientific theory of the world (*viz.*, physics) range over both physical objects and numbers, then, given his naturalism, Quine's physicalism embraces both concrete objects and abstract objects. He is a scientific realist regarding (observable and unobservable) physical objects and a Platonic realist regarding numbers (or sets). However, in *Pursuit of Truth* (1980) Quine downgrades the philosophical importance of ontology, including physicalism. He does so because of ontological relativity (*i.e.*, indeterminacy of reference). The thesis is that a theory's ontology can be supplanted *salva veritate* by any one-to-one mapping of it. Ontological relativity thus engenders an attitude of indifference toward various equally apt ontologies for a given theory, including physical theory so called. At the same time it highlights the importance of a theory's ideology, that is, its lexicon of predicates. The philosophical point of Quine's thesis is, then, that what a theory says there is is less important to our understanding of the world than what a theory says about what there is.

There are two further senses in which Quine may be said to be a physicalist. First, as expected, he rejects Cartesian dualism of mind and body in favor of materialism. In this regard, he endorses Donald Davidson's anomalous monism: token identity, type diversity. Second, he is a physicalist in the sense in which physicalism is opposed to phenomenalism in epistemology (see below).

**EPISTEMOLOGY: EMPIRICISM.** If the best current scientific theory (tentatively and fallibly) proffers a physicalist answer to the question of what there is, then what does it proffer in response to the question of how we know what there is? The answer is, in a word, empiricism. Quine maintains that it is a finding of science that all that we come to know about the world begins with the activation of our nerve endings.

So, Quine endorses the naturalization of both ontology and epistemology. And although he downgrades the philosophical importance of ontology, he maintains the philosophical importance of epistemology. The central question of epistemology, according to Quine, is How do we acquire our theory of the world and why does it work so well? Any answer to this question must explain the relation between one's empirical data (the "meager input") and one's theory of the world (the "torrential output"). Much of what Quine wrote after *Word and Object* is, ultimately, devoted to answering this question. His own distinctive answer may be called externalized empiricism in order to differentiate it from approaches of other naturalized epistemologists (*e.g.*, Donald Davidson). Quine's empiricism is externalized in the sense that he takes sets of activated nerve endings as his data and sets of sentences as his theory of the world (as opposed, say, to impressions and ideas, respectively).

In Quine's hands, the general relation,  $R_1$ , holding between sets of activated nerve endings and sets of sentences gets analyzed into two relations. There is the causal relation,  $R_2$ , holding between holophrastically construed observation sentences and their respective patterns of activated nerve endings, and there is the logical relation,  $R_3$ , holding between those same observation sentences, now analytically construed, and standing sentences. Quine schematizes how the child or the race, beginning with verbal responses conditioned to their respective patterns of nerve endings ( $R_2$ ), could have gone on to achieve verbal reference to bodies, substances, unobservables, and abstract objects ( $R_3$ ). Moreover, his account of  $R_3$  explains how observation sentences are logically related to theoretical sentences in such a way that no bridge principles are needed for linking observation and theoretic sentences. His account also highlights the hypothetico-deductive method of prediction and falsification and the moderately holistic character of theory revision.

**RECIPROCAL CONTAINMENT.** Externalized empiricism is Quine's contribution to answering the central epistemological question of how we acquire our theory of the world and why it works so well. As such, his epistemology (empiricism) "contains" his ontology (physicalism): *nihil in mente quod non prius in sensu*. However, Quine's epistemologizing always takes place within some accepted theory of the world (the best one he can muster at the time), so his epistemology (empiricism) is itself contained within his ontology (physicalism). This latter containment is the central lesson of naturalism: There is no first philosophy. It is this latter containment that also



makes Quine's epistemology such a radical departure from the tradition.

### CHANGES OF MIND

Even though Quine's thought has been remarkably consistent since his first works appeared in the 1930s, he changed his mind on a few important matters. First, he downgraded the importance of ontology, discussed above. Second, in the context of radical translation, Quine dropped the idea that the linguist can translate the native's "Gavagai" as her own "Lo, a rabbit" just in case the native's stimulus meaning for "Gavagai" is approximately the same as the linguist's for "Lo, a rabbit." The problem is with making scientific sense of this "implicit homology assumption" regarding different people's nerve endings. Quine changed to the position that the linguist can translate the native's "Gavagai" as her own "Lo, a rabbit" just in case the linguist can empathize with the native to the extent that she can confidently conjecture that, were she in the native's position when he uttered (or assented to) "Gavagai," then she would have done likewise for "Lo, a rabbit." In this way the linguist is (tentatively) equating the native's "Gavagai" with her own "Lo, a rabbit" without relying on an implicit homology assumption. Third, since, according to Quine's externalized empiricism, the meager input underdetermines the torrential output, then it is conceivable that there could be two (or more) global theories of the world that are empirically equivalent, logically compatible, equally simple, and so forth. Would both be true? Quine's empiricism encourages an ecumenical response: Both would be true. His naturalism encourages a sectarian response: Only one would be true. Quine himself vacillated on the issue but eventually endorsed the sectarian response. This suggests that his commitment to naturalism runs deeper than his commitment to empiricism.

**See also** Anomalous Monism; Davidson, Donald; Empiricism; Epistemology; Ethics; Logic, History of; Materialism; Metaethics; Metaphysics; Naturalism; Naturalized

Epistemology; Ontology; Phenomenalism; Philosophy of Language; Philosophy of Mind; Philosophy of Science, History of; Philosophy of Science, Problems of; Physicalism; Reference; Set Theory; Subject and Predicate.

### Bibliography

#### WORKS BY QUINE

Among Quine's post-*Word and Object* books that bear directly on the topics discussed here are:

*The Ways of Paradox and Other Essays*. New York: Random House, 1966; enlarged ed., Cambridge, MA: Harvard University Press, 1976.

*Ontological Relativity and Other Essays*. New York: Columbia University Press, 1969.

*Roots of Reference*. La Salle, IL: Open Court, 1974.

*Theories and Things*. Cambridge, MA: Harvard University Press, 1981.

Other of Quine's later books of interest are:

*Set Theory and Its Logic*. Cambridge, MA: Harvard University Press, 1963; rev. ed., 1969.

*Philosophy of Logic*. Englewood Cliffs, NJ: Prentice Hall, 1970; Cambridge, MA: Harvard University Press, 1986.

*The Web of Belief*. With J. S. Ullian. New York: Random House, 1970; rev. ed., 1978.

*The Time of My Life*. Cambridge, MA: MIT Press, 1985.

*Quiddities*. Cambridge, MA: Belknap Press of Harvard University Press, 1987.

#### WORKS ON QUINE

The secondary literature on Quine is immense and still growing. It includes:

Barrett, R., and R. Gibson, eds. *Perspectives on Quine*. Cambridge, MA: Blackwell, 1990.

Davidson, D., and J. Hintikka, eds. *Words and Objections*. Dordrecht: Reidel, 1969.

Gibson, R. F., Jr. *The Philosophy of W. V. Quine: An Expository Essay*. Tampa: University Presses of Florida, 1982.

Hahn, L. E., and P. A. Schilpp, eds. *The Philosophy of W. V. Quine*. La Salle, IL: Open Court, 1986.

Hookway, C. *Quine: Language, Experience, and Reality*. Stanford, CA: Stanford University Press, 1988.

Romanos, G. D. *Quine and Analytic Philosophy*. Cambridge, MA: MIT Press, 1983.

**R. F. Gibson (1996)**





## RACISM

“Racism” is the doctrine that one group of men is morally or mentally superior to another and that this superiority arises out of inherited biological differences. Of the modern theories aimed at dividing one portion of humanity from another, it is the most morally reprehensible and the least substantially based. Nationalism has a certain rationale in the existence of nation-states, and it does not, at least not necessarily, imply the inferiority of one nation to another. The various doctrines of the struggle between economic classes can point to a wide assortment of empirical evidence in support of their claims; in the Marxist version the exploiting capitalist is as much a victim of the capitalist system as is the exploited proletariat, and the eventual overcoming of all class distinctions is a moral aim as well as a prophesied event. The tenets of racism, however, lead to moral conclusions that contradict many of the most generally accepted civilized standards and have notoriously led to what on ordinary grounds are inconceivable crimes. It might be claimed that ordinary standards are mistaken and that, for example, it was morally imperative that the Nazis exterminate the Jews—if racist claims had a substantial factual basis. Fortunately for ordinary moral standards, if not for the

exterminated Jews and other victims of racial persecution, the tenets of racism are not merely unsubstantiated by the facts but in large measure contradicted by the facts.

Nor have the most important racist theorists been equipped to judge the alleged facts on which they based their claims. The question of race is an enormously complex one, and a judgment on it requires a synthesis of materials from history and prehistory and from a wide variety of biological, anthropological, and psychological disciplines, but primarily from genetics. Many of the necessary facts have only recently become available, and major questions remain unanswered. Yet most racist theories were put forth prior to the accumulation of this evidence, and even most contemporary racist theories are based on outdated biology. Furthermore, most racists—Houston Stewart Chamberlain, with his varied but erratic education, is a possible exception—have lacked the scientific training required to judge whatever evidence was available at the time they wrote. And until a racist theory can be substantiated to a very high degree of probability, the unsavoriness of the conclusion that there are inequalities in the capacities of groups of men requires that the theory be rejected.

## OUTLINE OF THE THEORY

Although there are many variations on the racist theme (the number of contradictions among racist claims, notably about which are the privileged races, is enough in itself to cast doubt on the tenability of the whole racist enterprise), a model set of racist tenets, divisible into three groups of claims, can be isolated.

The first group starts with the premise that humankind is now, has been in the past, or ought to be in the future divided up into biologically distinct groups. The different tenses must be distinguished because in some instances the claim is made that the superior race is not now in existence but should be bred from the “best blood” among various existing groups. This claim is the link between racism and eugenics, but although eugenicists often fall into racist language or hold racist beliefs (Sir Francis Galton, the founder of eugenics, rated blacks as about two grades below the Anglo-Saxon “race,” and the British pragmatist philosopher F. C. S. Schiller supported both eugenics and the English fascist Sir Oswald Mosley), the connection between the two theories is not inevitable.

The distinction between groups of humankind is held to be based on the common biological heredity of the members of each group. Among the biological distinctions between groups are inherited capacities for certain cultural activities—some races, it is claimed, are more warlike than others, some more musical, others predestined to be dominated. These are factual claims, seemingly open to confirmation or refutation by a scientific examination of the evidence, and the evidence seems overwhelmingly against every one of them. Someone who upheld these views would not necessarily be a racist, but they are essential to the racist position.

In a class by itself is the claim that the mechanism of transmission of group characteristics is the blood. Of all racist claims this is the one most surely refuted, and it would seem to be inessential to the doctrine. Yet the insistent stress on this claim even in the face of overwhelming evidence of its falsity is an index of the nonrational sources of racist thinking. Theories of inheritance through the blood, of blood kinship, of bluebloods, and of good and bad blood are survivals of age-old prescientific thought, on the same order as the view that the soul is the breath.

The final set of doctrines are essential to racism and distinctive of it. Not only are human groups different from one another but some are “better,” “stronger,” “higher,” or “more creative” than others—physically,

intellectually, or morally. (The proponent of a particular racist doctrine quite naturally almost always identifies himself with the race he judges superior. Thus, Comte Joseph Arthur de Gobineau, who was born in the south of France and who placed the “Nordic race” at the pinnacle of humanity, devoted considerable research to proving his own descent from the Viking Otto Jarl; the British-born, French and Swiss-educated Chamberlain, who espoused Aryan or Teutonic superiority, included the contemporary English—and the Slavs and Celts—among the Teutons.) The higher race or races, it is claimed, have a moral right to dominate, to enslave, or even to eradicate the lower races. Finally, higher and lower races should not intermarry. Race mixture, or “mongrelization,” is against nature. For the superior race it can lead only to the lowering of standards and to racial degeneration. It would seem that race mixture would improve the “lower” race, but this is generally denied either on biological or on historical grounds. Thus, Chamberlain held that the “lower” Jewish race was not improved by an alleged ancient admixture of Aryan blood, which came too little and too late.

## CRITICISM

No complete examination of the fallacies of racist doctrines can be presented here. What seems most important is that there are not now and, so far as anthropological evidence shows, have never been any pure races of men and that the very concept of race as applied to groups of human beings is suspect. In the vast number of its traits humankind is one, and there has been constant intermarriage and a consequent diffusion of genetic traits throughout the species. There are obvious dissimilarities among groups of people, but these differences more or less gradually shade off into one another; it is a question of statistical predominance of certain physical or physiological traits in a population rather than of sharply defined group differences.

Estimates of the number of genes in man range from 10,000 to 100,000, whereas the number of genes that control skin color, shape of lips and nose, and hair form are few. Racists want to correlate these obvious differences—which in themselves are purely statistical and thus no certain guide to the ancestry of a particular person—with differences in innate inheritable mental characteristics. Yet the evidence is against any such correlation. Each gene or gene cluster, except for certain linked genes, is inherited individually; on the average, half comes from the father and half from the mother. The number of possible combinations of ancestral genes is astronomical, and the

question whether specific mental characteristics are linked with a particular genetic heritage can almost certainly be answered in the negative for human beings, if not, perhaps, for certain domestic animals.

In any case, humankind has apparently been faced with an environment that puts a premium on intelligence, and there seem to be no detectable group differences in intelligence. It is practically impossible to devise a satisfactory test to determine whether there are biological differences in intelligence. In most cases the available methods of classifying by ancestry those to be tested are quite fallible. It is equally difficult to find two groups genetically distinct and culturally alike, and intelligence tests are quite generally distorted by cultural factors and place a premium on particular cultural achievements that obscures any possible genetic factors in the results.

Finally, if there were any evolutionary reasons for thinking that some race was at one time constitutionally better fitted to one environment than another, the rate of human cultural change is such that this supposed superiority would have been insignificant for many centuries. There is no reason to think that one group of humankind is mentally or physically better fitted than another to cope with the complexities of modern urban civilization and an internationally dispersed technology.

To the above summary and inadequate account of the biological claims that contradict racism should be added the overwhelming historical evidence of constant migrations and intermarriages of human groups and the highly probable inference that movement and mixture was also the rule during the prehistory of the human species. This has been especially true of the two alleged races most notoriously prominent in racist literature, Aryans and Jews. The Aryan is generally presented as a pure and superior race and the Jewish "race" as inferior, contradictorily characterized as both pure and bastardized, often by the same author. However, there neither is nor could be evidence that either race is more or less "pure" than the other. Each group is an amalgam of people of varied ancestry, and mixture has produced no apparent genetic debilitation of the sort that racists inveigh against when they deplore the "mongrelization of the race." Cultural differences exist between Germans and Jews, but there are likewise cultural differences between different groups of Germans and between groups of Jews, as well as cultural similarities between German and Jewish groups. To assign these likenesses and similarities to race rather than to a vast complex of recognized socio-cultural factors is to ignore a great bulk of historical evidence.

**THE IRRATIONALISM OF RACISTS.** Arguing with a proponent of racism is like arguing with someone who would today claim that the earth is flat and at the center of the universe. The evidence that the earth is round is so overwhelming, and so bound up with our very conception of what physical science is, that in the face of someone who claims that the earth is flat we can only point helplessly at the great body of scientific factual claims and scientific laws and ask, "But don't you see?" Similarly, when we are faced with the claims of a racist who persists in his doctrine in the face of our very notions of what constitutes biology and what constitutes historical research, we have no common ground for argument with him. An extreme but typical racist statement can be used as an example:

It is established for all time: "alien albumen" is the sperm of a man of alien race. The male sperm is partially or completely absorbed by the female and thus enters her bloodstream. One single cohabitation of a Jew with an Aryan woman is sufficient to poison her blood forever. Together with the "alien albumen" she has absorbed the alien soul. Never again will she be able to bear purely Aryan children ... they will all be bastards. (Julius Streicher, quoted in Quentin Reynolds, Ephraim Katz, and Zwy Aldouby, *Minister of Death*, New York, 1960, p. 150)

To someone with the most elementary acquaintance with contemporary biology it is unnecessary to point out the false assumptions and false statements in this quotation. But to refute the argument in a way that would satisfy its maker is impossible, because he denies the very grounds on which a scientific refutation as we understand it could be based.

The racist views in Adolf Hitler's *Mein Kampf* likewise seem based on a different biology from the one we know, but in Hitler's thought there is an added historical dimension. The picture Hitler draws of the sociopolitical situation in Germany and Austria during his own lifetime is often shrewd, but it is open to rational criticism: he makes factual claims that can be shown to be historically untrue and historical interpretations that can be challenged by an appeal to evidence and probability. His picture of the Aryans as the only culture-creating people, whose presence in a certain area at a certain time can be demonstrated simply because cultural innovation must have taken place then and there, bears no relation to what we know of the movements of peoples or to our notions of probability. In the chapter "Nation and Race," Hitler

uses few examples, and when examples are given they are used tendentiously to show what they could not prove. Thus, the culture of contemporary Japan, he claims, is the product of European stimulation: It is Western culture and technology with Japanese trimmings. Without continued infusions of Western culture, the culture of contemporary Japan is doomed to decay, and the culture found in Japan by Western explorers must itself have been the ossified remnants of some earlier, but forgotten, Aryan invasion. Hitler's arguments do not generally reach even the level of this one, circular as it is. Yet to show that no such invasion took place in historical times, and probably could not have taken place in prehistoric times, seems no answer to Hitler's claims. The picture he presents of the past is a deliberately mythical one, on a deliberately mythical time scale that bears no apparent relation to the known events and temporal ordering of history. In the absence of such relationships, all appeals to facts become irrelevant, and facts are notably absent from the argument.

#### RACISM OUTSIDE GERMANY

Although racism as a fully articulated doctrine and the central feature of official policy is notoriously associated with Germany, it has been powerful elsewhere. It was among thinkers of the French Enlightenment—Comte de Boulainvilliers, Comte de Buffon, and Baron de Montesquieu—that the concept of race was first made explicit and the germs of racism were implanted. Gobineau, in the mid-nineteenth century, was the true originator of the doctrine of racism, and throughout the nineteenth century and later, French thinkers vied with one another to show their descent from Gauls, Romans, Gallo-Romans, Celts, or Teutons and the superior Frenchness of one of these purported races over another.

In the United States and England also racism has flourished, and in these countries the complex interconnection of racist doctrines with social and economic factors is most apparent. In English thought racism has been mainly a concomitant of imperialism. The influx of darker-skinned peoples from the Commonwealth has led both to widespread resentment and to the expression of racist sentiments, but not as yet to any new fully developed racist theories. In the United States racism first arose in the South as a defense of slavery, was invoked as a justification of American imperialist expansion into the western Pacific and the Caribbean and for the restriction of the immigration of "undesirable" stock into the United States, and arose again as a defense of segregation.

Twentieth-century arguments that blacks were biologically inferior are not essentially different from earlier ones, of which Samuel Cartwright's "The Prognathous Species of Mankind" (1857) is an example. The argument moves from stressed and exaggerated physiological differences between blacks and whites to the claim of broad mental differences. Features of the "typical negro" are closer to "the simiadae and the brute creation" than to whites. The standard black color is a shiny, oily black, and lighter colors are the result not of intermixture with whites but of sickness or degeneration. In "the bleaching process of bad health or degeneration" even the pigment of the iris is lost, and the degenerate Negro is clairvoyant at night. The Negro does not have real hair: "the shaft of each hair is surrounded with a scaly covering like sheep's wool, and, like wool, is capable of being felted. True hair does not possess this property.... the negro approximates the lower animals in his sense of smell, and can detect snakes by that sense alone. All the senses are more acute, but less delicate and discriminating than the white man's." Natural history, like the Bible, "proves the existence of at least three distinct species of the genus man, differing in their instincts, form, habit, and color. The white species having qualities denied to the black—one with a free and the other with a servile mind—one a thinking and reflective being, the other a creature of feeling and imitation, almost void of reflective faculties, and consequently unable to provide for and take care of himself."

Several racial theories, notably those of Madison Grant and Lothrop Stoddard, reflected the growing awareness among the descendants of earlier groups of immigrants to the United States of the changing national origins of later groups. The works of these men both promoted the fear of the ultimate extinction of the "white race" (which was often meant to exclude southern and eastern Europeans) by rising birth rates among Asians and Africans and influenced the restrictive immigration laws of the 1920s. But it is doubtful whether these or later writers have added anything substantially new to the racist theses.

*See also* Affirmative Action; Boulainvilliers, Henri, Comte de; Buffon, Georges-Louis Leclerc, Comte de; Chamberlain, Houston Stewart; Enlightenment; Fascism; Gobineau, Comte Joseph Arthur de; Montesquieu, Baron de; Nationalism; Schiller, Ferdinand Canning Scott.

## Bibliography

### RACIST WRITINGS

The fountainhead of racist doctrines is Count Joseph Arthur de Gobineau's four-volume *Essai sur l'inégalité des races humaines* (Paris, 1853–1855); Vol. I has been translated by Adrian Collins as *Essay on the Inequality of Races* (London and New York, 1915). The first two volumes were also published at Philadelphia as early as 1856 in a translation by H. Hotz and with an introduction by Josiah C. Nott, both of whom were propagandists for slavery. For an early refutation of Gobineau's views based on moral grounds, see Alexis de Tocqueville's correspondence in his "The European Revolution" & *Correspondence with Gobineau*, edited by John Lukacs (Garden City, NY: Doubleday, 1959).

Richard Wagner published his anti-Semitic essay, "The Jews in Music," in 1850. Wagner later became an enthusiastic supporter of Gobineau, and Gobineau of Wagnerism. Representative writings of Wagner on race are available in *Wagner on Music and Drama*, compiled by Albert Goldman and Evert Sprinchorn (New York, 1964). Also important in disseminating Gobineau's views in Germany was Ludwig Schemann, founder of the Gobineau-Verein, translator, editor, and biographer of Gobineau, and author of such racist works as *Die Rassenfrage im Schrifttum der Neuzeit* (Munich, 1931).

Houston Stewart Chamberlain, Wagner's son-in-law, ranks with Gobineau as a race theorist; the two-volume *Die Grundlagen des neunzehnten Jahrhunderts* (Munich, 1899), his major work, was translated by John Lees as *The Foundations of the Nineteenth Century*, 2 vols. (London and New York: J. Lane, 1911). Chamberlain influenced both Adolf Hitler's *Mein Kampf* (2 vols., Munich, 1925–1927) and Alfred Rosenberg's *Der Mythus des 20. Jahrhunderts* (Munich, 1930).

The term *Aryan* was popularized by Friedrich Max Müller as a label for the speakers of the hypothetical language from which Indo-European languages were allegedly descended. Although Müller later denied that the term had any racial significance, the romantic claim that language expresses the soul of the race made the identification of Aryan speakers with an Aryan race almost inevitable, and Müller's own writings abound in such identifications. See, for example, *Lectures on the Science of Language*, 2 vols. (London: Longman, Green, Longman, and Roberts, 1861–1864).

In defense of slavery on racial grounds, see, in addition to Cartwright's essay, Josiah Nott's *Types of Mankind* (Philadelphia, 1854), parts of which are reprinted with Cartwright's essay and other writings in *Slavery Defended: The Views of the Old South*, edited by Eric L. McKittrick (Englewood Cliffs, NJ: Prentice-Hall, 1963). Other American works are Madison Grant, *The Passing of the Great Race* (New York: Scribners, 1916), and Lothrop Stoddard, *The Rising Tide of Color against White World-Supremacy* (New York: Scribners, 1920).

### WORKS ON RACE AND RACISM

Jacques Barzun, *Race: A Study in Superstition*, 2nd ed. (New York: Harper and Row, 1965), is a historical survey. See also Ernst Cassirer, *The Myth of the State* (New Haven, CT: Yale University Press, 1946), Ch. 16, and Hannah Arendt, *The*

*Origins of Totalitarianism*, 2nd ed. (New York: Meridian, 1958), especially Ch. 6.

Ashley Montague, *Man's Most Dangerous Myth*, 4th ed. (New York, 1964), and Ashley Montague, ed., *The Concept of Race* (New York: Free Press of Glencoe, 1964), together survey much of the present relevant biological knowledge and opinion. L. C. Dunn and Theodosius Dobzhansky, *Heredity, Race, and Society*, 2nd ed. (New York: New American Library, 1952), is a clear and useful account of the genetic aspects. Carlton S. Coon, *The Origin of Races* (New York: Knopf, 1962), is a work by a physical anthropologist who believes in the existence of biological differences between human groups that are associated with intellectual differences. Henry E. Garrett, "The Equalitarian Dogma," in *Perspectives in Biology and Medicine* 4 (1961): 480–484, presents a minority view by a former head of the American Psychological Association.

*Philip W. Cummings (1967)*

## RACISM [ADDENDUM]

Racism is the view that (1) the human species is composed of different racial groups, (2) these groups are arranged hierarchically from least to most superior, and (3) superior groups have the right to use inferior groups for the benefit of the superior group. Sexism and speciesism are similarly defined, justifying the right of men to dominate women, and the right of human beings to dominate other species of beings. Before Darwin, racial groups were typically defined in terms of lineage and type: a racial group was the progeny of certain original types, each of which exemplified a distinctive physiology and pattern of behavior. The current status of a racial group was then explained by reference to its ancestral sources. Thus, the democratic and enterprising nature of the English was the result of their Anglo-Saxon heritage, whereas the servile position of Africans was the result of their being the progeny of Ham, cursed by Noah to be servants of servants. Some argued that Africans were not the progeny of Adam and Eve, but were a pre-Adamite lower species to be used for human benefit.

Post-Darwinian biology favored the notion of race as a sub-species—a group within a particular species that is isolated genetically from other members of that species, and as a result develops distinctive morphological and/or behavioral attributes. Africans, Asians, and Europeans look different because they have evolved on different continents and have developed different body types and personalities.

Social Darwinism portrayed evolution as a struggle for existence in which superior races survived and inferior ones perished. Eugenicists hoped to enhance natural

selection by using our knowledge of natural phenomena to reproduce superior human beings and avoid reproducing inferior ones. Polygenecists considered the “lower” races to be of a different species than the “higher” races, and therefore subject to the interests of the higher races. They believed that, like the offspring of horses and donkeys, racial “hybrids” were likely to be infertile and dysfunctional. Even when races were acknowledged to be of the same species, “race mixing” was considered dysgenic and debilitating to the “higher” races. Miscegenation laws prohibiting intermarriage and procreation between different races were considered to be in society’s best interests and therefore good social policy.

The claim that Africans and African Americans have diminished cognitive and moral capacities compared with Europeans and European Americans was often used to justify slavery and segregation. This claim continues to attract adherents who seek naturalistic explanations of skewed racial achievements such as J. Richard Herrnstein and Charles Murray (1994), Michael Levin (1997), and Stephen Kershner (2003). According to Allan Chase (1980), Malthusians held that the diminished intelligence of the “lower races” lead them to bad choices and immoral behavior that worsened their plight and made them among the least well off wherever they were.

Some have argued that research on racial differences perpetuates harm by reinforcing racist assumptions and should be curtailed. But while the results of research on racial differences could be used to harm, such knowledge could also be used to help. If, for instance, it were found that people with high melanin content in their skin responded to a particular chemical compound that affected mental functioning, then it might be possible to manipulate that compound to either boost or retard intellectual performance. Research on racial differences degenerates into racism only when racial differences are believed to establish a hierarchical ranking that is biologically fixed and immutable.

Many now consider the very concept of a race to be an artifact of European expansionism, justifying European domination of African, Asian, and Native American people. On this view, Europeans classified “others” as different races to further the ends of domination, and continued use of racial categories merely reinforces that original aim. Rejecting European economic, political, and cultural imperialism requires that we reject both racism and racialism (i.e., classification by races). In a similar fashion, Marxists consider racism to be an ideological ploy that divides the lower classes so that the European

and non-European proletariat fight one another instead of fighting capitalists.

The Nazis portrayed Jews as an inherently acquisitive and parasitical race that threatened the evolution of mankind. Their mission was to exterminate the “Jewish race” and establish the unchallenged hegemony of the “Aryan race.” In reaction to the atrocities of the holocaust, many scientists and leaders such as Joseph Graves (2001) and Ashley Montagu (1997) marshaled evidence to show that race was a pseudoconcept with no biological validity. On this view, neither the Jews nor the Aryans were races defined by distinct biological differences. Jews typically were biologically more similar to contiguous non-Jews than they were to Jews in distant locales. Moreover, classifying people into races by skin color, hair texture, and other observable characteristics ignores many other features (such as internal proteins and DNA sequences) that could also be used to classify them. Those who argue that there are no races typically cite evidence that genetic variation is greater within traditional racial groups than between them, thus showing that racial groups are not reproductively isolated gene pools.

But other biologists and social scientists, such as Phillip Kitchner (1999), Robin Andreasen (2000), and Neil Risch (2003), have insisted that there is compelling biological evidence for the existence of races. These researchers assert that when human populations are classified in terms of their ancestral geographic origins, there is a high correlation between traditional racial groups and genetic clusters.

However, the question of whether or not races exist is independent of whether racism exists. Just as it is possible for witchcraft to exist even though there are no witches, so it is possible for racism to exist even if there are no races. Some, such as Naomi Zack (2002), argue that if we are to move beyond “racial” animosities of the past, we must cease using racial categories, because they have no biological validity. For Lucius Outlaw (1996) and Alain Locke (1999), continued consciousness of racial distinctions may be linked to pride in cultural achievements made under extreme duress, or to demands for restitution for past and present harms. Whether or not races exist, it should be possible to agree that racism is morally wrong, and that racism should be eliminated.

Racist behavior elevates the interests of members of allegedly superior races over the interests of allegedly inferior races. The existence of racist behavior may be independent of individual intent, as is often the case with institutional racism, where certain procedures and practices harm groups historically considered inferior, even if



the implementers do not explicitly intend such harm. Thus, requirements that are unnecessary for successful performance and recruitment limited to traditional networks often serve to perpetuate the effects of overt and egregious racist acts of the past, even if this is not currently intended by those who implement such policies.

**See also** Civil Disobedience; Cosmopolitanism; Multiculturalism; Postcolonialism; Republicanism; Social and Political Philosophy.

### Bibliography

- Andreasen, Robin. "Race: Biological Reality or Social Construct" *Philosophy of Science* 67 (Proceedings) (2000): 653–666.
- Appiah, Kwame Anthony. *In My Fathers House: Africa in the Philosophy of Culture*. New York: Oxford University Press, 1992.
- Banton, Michael. *Racial Theories*. New York: Cambridge University Press 1987.
- Block, N. J., and Gerald Dworkin, eds. *The I.Q. Controversy: Critical Readings*. New York: Pantheon, 1976.
- Braun, Lundy. "Race, Ethnicity, and Health: Can Genetics Explain Disparities?" *Perspectives in Biology and Medicine* 45 (2002): 159–174.
- Chase, Allan. *The Legacy of Malthus: The Social Costs of the New Scientific Racism*. Urbana: University of Illinois Press, 1980.
- Ezorsky, Gertrude. *Racism and Justice: The Case for Affirmative Action*. Ithaca, NY: Cornell University Press, 1991.
- Frederickson, George M. *Racism: A Short History*. Princeton, NJ: Princeton University Press, 2002.
- Graves, Joseph L. Jr. *The Emperor's New Clothes—Biological Theories of Race at the Millennium*. New Brunswick, NJ: Rutgers University Press, 2001.
- Harris, Leonard, ed. *Racism*. Amherst, NY: Prometheus Books, 1999.
- Herrnstein, Richard, and Charles Murray. *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press, 1994.
- Kershner, Stephen. "Experiential Diversity and Grutter." *Public Affairs Quarterly* 17 (2) (April 2003): 163–164.
- Kitchner, Phillip. "Race, Ethnicity, Biology, Culture." In Harris, *Racism*, 87–117.
- Levin, Michael. *Why Race Matters: Race Differences and What They Mean*. Westport, CT: Praeger, 1997.
- Locke, Alain. "The Contribution of Race to Culture." In Harris, *Racism*, 215–218. Amherst, NY: Humanity, 1999.
- Montagu, Ashley. *Man's Most Dangerous Myth: The Fallacy of Race*. Walnut Creek, CA: Altamira Press, 1997.
- Mosley, Albert. "Are Racial Categories Racist?" In *Research in Africana Literatures*, 28.4 (Winter 1997), pp. 101–110.
- Outlaw, Lucius T. Jr. *On Race and Philosophy*. New York: Routledge, 1996.
- Risch, Neil, Esteban Burchard, Elad Ziv, and Hua Tang. "Categorization of Humans in Biomedical Research: Genes, Race, and Disease." *Genome Biology* 3 (2002): comment 2007.1–2007.12.

Zack, Naomi. *Philosophy of Science and Race*. New York: Routledge, 2002.

*Albert G. Mosley (1996, 2005)*

## RADBRUCH, GUSTAV

(1878–1949)

Gustav Radbruch was a German legal philosopher whose name and work have become widely known outside Germany only since the end of World War II. During his lifetime, the interests and activities of scholar, politician, and reformer of law were closely intermingled. After World War I, Radbruch became active in the Social Democratic Party and twice served as minister of justice of the Weimar Republic. His principal work was the draft of a new criminal code. Later he held a chair of law at the University of Heidelberg, from which he was dismissed by the Nazi regime. After World War II he was recalled and exercised a predominant influence in the reorientation of German legal education and philosophy until his death.

Radbruch's legal philosophy, generally known as "relativism," is closely akin to the position of his friend and teacher Max Weber. Radbruch believed, like Weber, that values could not be scientifically proved and that they were "a matter of conscience (*Gewissen*), not of science (*Wissenschaft*)," This in no way implied indifference to values. Radbruch differed both from Rudolf Stammler, who sought to formulate a theoretically valid concept of justice, and from Hans Kelsen, who detached legal science altogether from a philosophy of values. Radbruch, while starting from the Kantian distinction of "is" (*Sein*) and "ought" (*Sollen*), was guided mainly by the teachings of Heinrich Rickert and Emil Lask in treating law as a *Kulturwissenschaft*, a science directed to the realization of values. He therefore considered that the task of legal philosophy was to relate legal reality to basic ideas. But the truth of specific ideas and values cannot be scientifically proved. Radbruch instead developed—and applied to numerous specific problems of law—a series of antinomies of legal values. Thus, the Aristotelian idea of distributive justice, which directs equals to be treated equally, says nothing about the perspective from which they are to be characterized as equals or unequals. Justice, which cannot yield objective criteria of equality, must be supplemented by a second value, "utility," and a third, "security." Between these three values there is constant tension. In another perspective, law can be directed to individual values, collective values, or work values. Accordingly, a legal system emphasizes either individualism, collec-

tivism, or transpersonalism. For the first, the ultimate idea is liberty; for the second, the nation; and for the third, civilization.

After the war, Radbruch recoiled from the extremes of tolerance—as practiced by the Weimar Republic during the rise of the Nazi movement—having witnessed the unprecedented barbarism of the Third Reich, which was largely covered by a formal notion of law. He tentatively turned to a moderate natural-law philosophy, holding that in certain extreme cases a contradiction between positive law and justice might reach such an intolerable degree that the law as unjust law (unlawful law, *unrechtes Recht*) must cede to the higher demands of justice. Radbruch died before he could elaborate his thesis beyond the postulate that special courts should be empowered to adjudicate the validity of laws.

**See also** German Philosophy; Justice; Kelsen, Hans; Philosophy of Law, History of; Philosophy of Law, Problems of; Rickert, Heinrich; Stammler, Rudolf; Value and Valuation; Weber, Max.

### Bibliography

#### WORKS BY RADBRUCH

- Einführung in die Rechtswissenschaft*. Leipzig, 1907. Edited by K. Zweigert. Stuttgart: Koehler, 1961.
- Rechtsphilosophie*. Leipzig, 1914. 3rd ed. revised and enlarged, Leipzig: Quelle and Meyer, 1932. 5th ed., completed with a preface by Erik Wolf, Stuttgart: Koehler, 1956. Translated from the German edition by Kurt Wilk with introduction by E. W. Patterson in *The Legal Philosophies of Lask, Radbruch and Dabin*. Cambridge, MA: Harvard University Press, 1950.
- “Gesetzliches Unrecht und übergesetzliches Recht.” *Süddeutsche Juristenzeitung* 5 (1946): 107ff.
- Der Geist des englischen Rechts*. 2nd ed. Heidelberg, 1947.
- Vorschule der Rechtsphilosophie*. Willsbach: Scherer, 1947.

#### WORKS ON RADBRUCH

- Campbell, A. H. *Gustav Radbruchs Rechtsphilosophie*. Hanover, 1949.
- Cattaneo, M. A. “Il Positivismo giuridico e la separazione tra il Diritto e la Morale.” *Rendiconti (Istituto Lombardo)* 94 (1960): 701–742.
- Friedmann, Wolfgang. “Gustav Radbruch.” *Vanderbilt Law Review* 14 (1960): 191–209.
- Wolf, Erik. *Griechisches Rechtsdenken*, 3 vols. Tübingen, 1950–1954.
- Wolf, Erik. *Grosse Rechtsdenker der deutschen Geistesgeschichte*. 3rd ed. Tübingen: Mohr, 1951.
- Wolf, Erik. *Das Problem der Naturrechtslehre: Versuch einer Orientierung*. Tübingen, 1955. 2nd ed. Tübingen, 1959.
- Wolf, Erik. *Rechtsgedanke und biblische Weisung*. Tübingen, 1948.

Wolfgang Friedmann (1967)

## RADISHCHEV, ALEKSANDR NIKOLAEVICH

(1749–1802)

Aleksandr Nikolaevich Radishchev was the leading social critic and philosopher of the Russian Enlightenment. He was born in Moscow, the son of a prosperous landowner, and was educated in Moscow, in St. Petersburg, and, from 1766 to 1771, at the University of Leipzig. At Leipzig he studied under the Leibnizian Ernst Platner and read widely in current French philosophy. Upon his return to Russia he pursued a successful career in the civil and military service until 1790, when his radical work *Puteshestvie iz Peterburga v Moskvu* (St. Petersburg, 1790; translated as *A Journey from St. Petersburg to Moscow*) aroused the ire of Catherine the Great and he was exiled to Siberia. Paul I permitted him to return to European Russia in 1796. After the accession of Alexander I, in 1801, Radishchev was appointed to a special legislative commission, but his egalitarian, libertarian proposals went unheeded, and in September 1802 he took his own life in St. Petersburg.

In the *Journey*, Radishchev employed the principles of natural law and the social contract to support a severe critique of Russian social institutions, serfdom in particular. Under the inspiration of Jean-Jacques Rousseau, Voltaire, Guillaume-Thomas-François de Raynal, and other French thinkers, he condemned serfdom as morally wrong and economically inefficient, criticized autocracy, and attacked censorship and other practices that violate men’s natural rights to freedom and equality. He advocated immediate reforms to avert revolution and called generally for enlightenment and “naturalness” in social arrangements, manners, and morals.

In Siberia, Radishchev wrote his principal philosophic work, *O cheloveke, o ego smertnosti i bessmertii* (On man, his mortality and immortality; published posthumously, St. Petersburg, 1809), a close examination of the cases for and against personal immortality. In the end he rejected materialistic denials of immortality in favor of various arguments—from personal identity and the conservation of force, among others—that suggest the existence of an incorporeal soul that survives the body and passes into a more perfect state. In epistemology Radishchev adopted a realistic position and accepted experience as the only basis for knowledge but maintained that in addition to sensory experience there is “rational experience” of the relationships of things and that man “feels” the existence of a Supreme Being. He also maintained that things in themselves are unknowable,

asserting that thought, like the verbal expression it employs, is merely symbolic of reality.

Radishchev's treatise *O cheloveke* was one of the first original philosophic works in the Russian language, and the influence his pioneering social criticism had on Alexander Pushkin, the Decembrists, and subsequent generations of Russian reformers and revolutionaries has led to his being regarded as the father of social radicalism in Russia. He was also a poet of considerable talent.

**See also** Immortality; Natural Law; Rousseau, Jean-Jacques; Russian Philosophy; Social Contract; Voltaire, François-Marie Arouet de.

### Bibliography

#### WORKS BY RADISHCHEV

*Polnoe sobranie sochinenii* (Complete works). 3 vols. Moscow and Leningrad, 1938–1952.

*A Journey from St. Petersburg to Moscow*. Translated by Leo Wiener. Cambridge, MA: Harvard University Press, 1958.

"On Man, His Mortality and Immortality." Selections translated by Frank Y. Gladney and George L. Kline in *Russian Philosophy*, edited by J. M. Edie, J. P. Scanlan, and M.-B. Zeldin. Chicago: Quadrangle, 1965.

#### WORKS ON RADISHCHEV

Dynnik, M. A. et al., eds. *Istoriia filosofii*. 6 vols. Moscow, 1957–.

Lapshin, I. I. *Filosofskie vzgliady A. N. Radishcheva* (Philosophical views of A. N. Radishchev). Petrograd, 1922.

Zenkovsky, V. V. *Istoriia russkoi filosofii*. 2 vols. Paris, 1948–1950. Translated by George L. Kline as *A History of Russian Philosophy*. 2 vols. New York: Columbia University Press, 1953.

**James P. Scanlan (1967)**

## RĂDULESCU-MOTRU, CONSTANTIN

(1868–1954)

Constantin Rădulescu-Motru, the Romanian philosopher of energetic personalism, studied at the universities of Bucharest, Paris, Munich, and Leipzig. He obtained his doctorate from Leipzig in 1893 with a thesis on the development of Immanuel Kant's theory of causality in nature, published in Wilhelm Wundt's *Philosophische Studien*. In 1904 he became professor of psychology and logic at the University of Bucharest. He founded the journals *Noua Revista Română* (1900), *Studii Filosofice* (1905, after 1920 called *Revista de Filosofie*), *Anale de Psihologie* (1935), and

*Jurnal de Psihotehnică* (1937), as well as the Romanian Society of Philosophy.

Rădulescu-Motru was the dominant figure in Romanian philosophy from 1905 to 1930. The most articulate expression of his philosophical system is to be found in his *Personalismul Energetic* (1927). Influenced by the work of Wilhelm Ostwald and William Stern, it was an impressive effort to unify the results of natural science, biology, and psychology. Rădulescu-Motru called his system personalistic because the human personality plays the central role within it, and energetic because he considered personality to be the highest form of cosmic energy. The universe is in continuous evolution, and its goal is the creation of energetic personality. Rădulescu-Motru distinguished six stages of the evolutionary process: cosmic energy, adaptation, organic individuality, consciousness, ego, and personality. Personality is both modified and enriched through evolution from the primitive *homo divinans* to *homo faber*. Finally, through Stoicism, Christianity, and science, the energetic personality, the vocational or professional man, emerges. With the achievement of a personality having a total comprehension of the universe, the evolutionary process will come to an end; Nature will have reached its ultimate goal.

**See also** Kant, Immanuel; Ostwald, Wilhelm; Personalism; Stern, Louis William; Wundt, Wilhelm.

### Bibliography

#### WORKS BY RĂDULESCU-MOTRU

*Puterea Sufleteasca*. Bucharest, 1907.

*Elemente de Metafizică* (1912). Definitive edition Bucharest, 1928.

*Curs de Psihologie*. Bucharest, 1924.

*Personalismul Energetic*. Bucharest, 1927.

*Vocația*. Bucharest, 1932.

*Românismul*. Bucharest, 1936.

#### WORKS ON RĂDULESCU-MOTRU

*Omagiu profesorului C. Rădulescu-Motru*. Bucharest, 1932.

Rus, G. "Il personalismo energetico di C. Rădulescu-Motru." *Acta Philosophica et Theologica* 2 (1964): 411–438.

**Mircea Eliade (1967)**

## RAHNER, KARL

(1904–1984)

One of the most significant Roman Catholic theologians of the twentieth century and a formative influence upon

Vatican II, Karl Rahner was born on March 5, 1904, in the city of Freiburg im Breisgau, Germany, the fourth of seven children in the family of Karl and Luise (Trescher) Rahner. Upon graduation from secondary school at the age of eighteen, Rahner followed in the footsteps of his elder brother Hugo and entered the Society of Jesus; he was to remain a Jesuit his entire life. During his novitiate studies from 1924 to 1927, Rahner was introduced to Catholic scholastic philosophy and to the modern German philosophers. He seems especially to have been influenced by the work of Joseph Maréchal (1878–1944), the Belgian philosopher and Jesuit, whose adoption of Kant's transcendental method in his five-volume work, *Le point de départ de la métaphysique*, had led to somewhat of a breakthrough in the appreciation of Kant's philosophy among neo-Scholastics. Maréchal was known as the "father of transcendental Thomism" for his use of St. Thomas Aquinas's epistemology in an attempt to demonstrate that the metaphysical world Kant had secured for practical reason was already inherent in the theoretical.

After teaching Latin at the Feldkirch Novitiate, Rahner studied theology at Valkenburg in the Netherlands (1929–1933), where his Christian spirituality was further nurtured through study of patristic and medieval mysticism, and above all of St. Ignatius of Loyola (1491–1556), founder of the Jesuit order and author of the *Spiritual Exercises*. Following ordination to the priesthood in 1932, Rahner commenced study for his doctoral thesis in philosophy at Freiburg, while at the same time attending lectures by Martin Heidegger, whose philosophy of Dasein, or "being in the world," was to be the other primary philosophical influence upon him. His dissertation, a response to Kant's critique of theoretical metaphysics by means of the transcendental Thomism of Maréchal and the existentialism of Heidegger, was rejected by his doctoral director, Martin Honecker, for its departure from more traditional neo-Scholastic interpretations of Aquinas's epistemology, but was later published as *Geist in Welt* (Spirit in the world).

After failing the doctorate in philosophy, Rahner returned to Austria, where he successfully completed his second dissertation, this time in theology, at Innsbruck in 1936 and was appointed as Privatdozent (lecturer) in the faculty of theology of the University of Innsbruck in 1937. That summer he delivered a series of lectures to the Salzburg summer school on the "Foundations of a Philosophy of Religion," later published as *Hörer des Wortes* (Hearer[s] of the word).

When the Nazis abolished the theology faculty (July 1938) and the Jesuit college (October 1939) at Innsbruck,

Rahner left for Vienna, where he did some teaching and served as a consultant at the Pastoral Institute for five years. After a brief stint as a pastor in Bavaria in the final year of the war, he taught dogmatic theology at Berchmanskolleg in Pullach. In 1948 he returned to the theology faculty at Innsbruck, where he was to reside until 1964. There he lectured on a wide variety of topics to later be included in the essays published as *Schriften zur Theologie* (Theological investigations), the first volume of which appeared in 1954. Of particular significance was his scholarly preoccupation with the relationship between nature and grace.

During this prolific period Rahner experienced some difficulties within the Church, beginning as early as 1950 when he was prevented from publishing a book on the Assumption of Mary, and continuing through the following decade until 1962, when he was placed under a censorship regulation from Rome. Suspicions over his orthodoxy subsided, however, when the newly elected Pope John XXIII appointed Rahner as one of the theological experts (*periti*) at the Second Vatican Council, and the censorship upon him was reversed in 1963. Rahner's influence at Vatican II was widespread; particularly noteworthy is his selection as one of the seven theologians who would develop *Lumen Gentium* (Dogmatic constitution on the Church), a document fully explicating the doctrine of the Church, and setting forth explicitly in chapter II the Church's inclusivist stance with regard to salvation.

It was during the Second Vatican Council that Rahner was invited to take the Chair in Christianity and Philosophy of Religion at the University of Munich, where he began teaching in 1964, the same year that a Festschrift, *Gott in Welt* (God in the world) was published in honor of his sixtieth birthday. During his time at Munich, Rahner published a collection of essays in spirituality as the seventh volume of *Schriften zur Theologie*, and together with Edward Schillebeeckx edited the first issue of *Concilium*. In 1967 Rahner accepted the University of Münster's invitation to become Ordinary Professor of Dogmatics and the History of Dogma, where he completed three more volumes of the *Schriften*, before retiring in 1971. Retirement brought him back to Munich, where he prepared *Grundkurs des Glaubens* (Foundations of Christian faith), the most systematic summary of his theology, and to Innsbruck, where in addition to pastoral and moral essays, Rahner worked out the most developed form of his transcendental Christology, and completed the final volumes of the *Schriften*, thus continuing the life

of the diligent scholar until his death in Innsbruck on March 30, 1984.

Rahner has been criticized for a failure to adequately address the problem of evil, especially in light of his experience as a German Catholic living through the Nazi genocide of two-thirds of the Jewish population of Europe. Neither in the above nor in what follows will it be possible to do full justice to the breadth of Rahner's theological output, which covers almost every aspect of religious thought. The focus here is upon Rahner's efforts at *aggiornamento*, or renewal of neo-Scholasticism and the philosophical import of two concepts integral to his theological *weltanschauung*: *Vorgriff auf esse* and *das übernatürliche Existential*.

*Geist in Welt* focuses upon one of the central problems of philosophy, namely the nature and possibility of metaphysics. In this work, Rahner examines one part of St. Thomas Aquinas's metaphysics of knowledge, specifically that section of the *Summa theologiae* that addresses what appears to sense intuition, *conversio ad phantasmata* (conversion to the phantasm), in light of Kant's critique of speculative metaphysics. Whereas Kant had rejected theoretical knowledge of God in order to secure a place for metaphysics as a practical philosophy, Rahner uses the tools of transcendental and existentialist philosophy, honed through Maréchal and Heidegger respectively, to retrieve the theoretical metaphysics of St. Thomas. *Spirit in the World* is thus Rahner's attempt to demonstrate how, given that human knowledge is wedded to the *a posteriori*, or realm of sensory experience, metaphysics is still possible; and as fraught as the philosophical analysis is, the main arguments are fairly accessible.

Essentially Rahner proffers a teleology of knowledge according to which there is presupposed in every human act of knowing the *Vorgriff auf esse* (the "pre-apprehension of being" in Heideggerian terms), a transcendental awareness of infinite being, or of God, the *a priori* condition without which no individual act of knowing could occur. In every act of knowing then, the individual, or "spirit in the world," has already reached out beyond the world and known the metaphysical. This awareness of God, which is always indirect and shrouded in mystery (since we cannot know God as if God were an object among other realities that are present to us), presupposes the transcendental orientation of the human knower to God, who is both the source and ultimate goal of the human quest for knowledge. To be human is therefore to be in relation to God, since we implicitly affirm the existence of God in every judgment we make, regardless of whether or not we ever formally acknowledge this. *Ipsa*

*facto*, human existence itself implies the transcendental experience of God for Rahner, thus satisfying not only the transcendental Thomism of Maréchal and the existentialism of Heidegger, but also the Ignation impulse to "find God in all things."

*Hörer des Wortes* is formally an investigation into the relationship between philosophy of religion and theology. Philosophy of religion, according to Rahner, consists in showing human beings to be the infinite spirits who, because of our nature, are turned toward a possible revelation, or self-communication of God, since revelation for Rahner is always personal, not propositional. God, the personal infinite, chooses human history as the place of transcendent self-communication (a divine self-communication that finds concrete historical expression in Jesus Christ); theology begins with the human person who has become attuned to God's self-communication, a hearer of God's word. In order to make this case, Rahner develops his "transcendental arguments" further and grounds them more fully theologically. By means of a "theological anthropology," a metaphysical analysis of human nature, Rahner proposes *Vorgriff auf esse* as a pre-apprehension of infinite being that also elicits the restless yearning of the human spirit (echoing a desire at least as old as Augustine) for fulfillment in and through that absolute being whose self-communication is both the ground and telos of human existence.

This understanding of the human spirit's desire for transcendent meaning, together with God's ineffable self-communication, later has important implications for Rahner's interpretation of the relationship between nature and grace, a relationship examined through his concept of *das übernatürliche Existential*, the "supernatural existential," first coined during his intervention in the *nouvelle théologie* debate in 1950 but still being worked out as late as 1976 in *Grundkurs des Glaubens*. The debate revolved around whether or not the human orientation toward God was natural or supernatural, and Rahner uses the term "supernatural existential" in an attempt to overcome the tendency in neo-Scholastic theology to dichotomize nature and grace, while at the same time safeguarding the gratuity of God's grace. Once again borrowing from Heidegger's vocabulary, Rahner defines an "existential" as a fundamental element in human existence, and claims that the central and abiding existential of human nature is the unconditional desire for grace and for the beatific vision. At the same time, however, he argues that the very fact of this desire already belies God's self-communication, precisely the meaning of grace for Rahner. In other words, since our very existence is per-

meated with God's constant self-giving, human nature is already grace laden. *Das übernatürliche Existential* ultimately entails the universal human experience of grace, and similar to the *Vorgriff auf esse*, this experience, though wedded to the world, is also transcendental and thus can never be directly or concretely realized.

The ubiquitous nature of the "supernatural existential" also undergirds Rahner's Christian inclusivism, itself a corollary of his philosophy of grace, and arguably the theological stance for which he is best known in non-Catholic circles. For Rahner, because God's gracious self-communication has found concrete historical expression in Jesus Christ, all grace is ultimately the grace of Christ; yet, significantly, Christ's universal grace is not narrowly circumscribed by Christianity. If *das übernatürliche Existential* is a universal given, and just as the *Vorgriff auf esse* is never directly or concretely realized, then it is possible that a person may accept this gift of grace without explicit acknowledgment and regardless of whether or not one is formally Christian. It is for this reason that non-Christians living lives of grace are "anonymous Christians" from Rahner's perspective, a title not intended as a subtle form of Christian supersessionism, but rather as a theologically astute commitment to the view that God's grace is active well beyond the confines of Christianity. The religious inclusivism espoused by Rahner had an ecumenical import that has proven vital to Catholic interreligious dialogue in the post-Vatican II era and that presumably will continue to be relevant to the burgeoning interest in religious diversity among philosophers of religion well into the third millennium

**See also** Heidegger, Martin; Maréchal, Joseph; Thomas Aquinas, St.

## Bibliography

### WORKS BY RAHNER

- Geist in Welt: Zur Metaphysik der endlichen Erkenntnis bei Thomas von Aquin.* Innsbruck and Leipzig: Felizian Rauch, 1939; 2nd ed., expanded and reworked by J. B. Metz. Munich: Kösel, 1957. Translated as *Spirit in the World* by W. Dych. New York: Herder and Herder, 1968.
- Grundkurs des Glaubens: Einführung in den Begriff des Christentums.* Freiburg, Basel, Vienna: Herder, 1976. Translated as *Foundations of Christian Faith: An Introduction to the Idea of Christianity* by W. Dych. London: Darton, Longman and Todd, 1978.
- Hörer des Wortes: Zur Grundlegung einer Religionsphilosophie.* Munich: Kösel-Pustet, 1941; 2nd ed., reworked by J. B. Metz. Munich: Kösel, 1963. Translated as *Hearer of the Word* by J. Donceel. New York: Continuum, 1994.

*Karl Rahner: Sämtliche Werke.* 32 vols., edited by K Lehmann, J. B. Metz, K. H. Neufeld, A. Raffelt, and H. Vorgrimler. Freiburg: Herder, 1995ff.

*Schriften zur Theologie*, 16 vols. Einsiedeln und Zurich, 1954–1984. Translated as *Theological Investigations*, 23 vols. London: Darton, Longman and Todd, 1961–1984.

### WORKS ABOUT RAHNER

- Bleistein, R., and E. Klinger. *Bibliographie Karl Rahner, 1924–1969.* Freiburg: Herder, 1969.
- Bleistein, R., ed. *Bibliographie Karl Rahner, 1969–1974.* Freiburg: Herder, 1969.
- Dych, W. *Karl Rahner.* London: Geoffrey Chapman, 1992.
- Imhof, P., and H. Treziak. "Bibliographie Karl Rahner, 1974–1979." In *Wagnis Theologie: Erfahrungen mit der Theologie Karl Rahners*, edited by H. Vorgrimler, 579–597. Freiburg: Herder, 1979.
- Imhof, P., and E. Meuser. "Bibliographie Karl Rahner, 1979–1984." In *Glaube im Prozess*, edited by E. Klinger and K. Wittstadt, 854–888. Freiburg: Herder, 1984.
- Kilby, K. *Karl Rahner: Theology and Philosophy.* London: Routledge, 2004.
- Marmion, D., and M. E. Hines, eds. *The Cambridge Companion to Karl Rahner.* Cambridge, U.K.: Cambridge University Press, 2005.
- Pedley, C. J. "An English Bibliographical Aid to Karl Rahner." *The Heythrop Journal* 25 (1984): 319–365.
- Raffelt, A. "Karl Rahner: Bibliographie der Sekundärliteratur, 1948–1978." In *Wagnis Theologie: Erfahrung mit dem Theologie Karl Rahners*, edited by H. Vorgrimler, 598–622. Freiburg: Herder, 1979.
- Raffelt, A. "Karl Rahner: Bibliographie der Sekundärliteratur, 1979–1983 und Nachträge." In *Glaube im Prozess*, edited by E. Klinger and K. Wittstadt, 872–885. Freiburg: Herder, 1984.
- Raffelt, A., and R. Siebenrock. "Karl Rahner: Sekundärliteratur, 1984–1993." In *Karl Rahner in Erinnerung*, edited by A. Raffelt, 165–205. Freiburger Akademieschriften, Band 8. Düsseldorf: Patmos, 1994.
- Tallon, A. "Rahner Bibliography Supplement." *Theology Digest* 38 (2) (1991): 131–140.
- Tallon, A. "Rahner Studies, 1939–1989: Part 1, 1939–1973." *Theology Digest* 36 (4) (1989): 321–346.
- Tallon, A. "Rahner Studies, 1939–1989: Part 2, 1974–1989." *Theology Digest* 37 (1) (1990): 17–41.

*Elizabeth Galbraith (2005)*

## RAMSEY, FRANK PLUMPTON (1903–1930)

Frank Plumpton Ramsey, the Cambridge mathematician and philosopher, was one of the most brilliant men of his generation; his highly original papers on the foundations of mathematics, the nature of scientific theory, probability, and epistemology are still widely studied. He also

wrote two studies in economics, the second of which was described by J. M. Keynes as “one of the most remarkable contributions to mathematical economics ever made.” Ramsey’s earlier work led to radical criticisms of A. N. Whitehead and Bertrand Russell’s *Principia Mathematica*, some of which were incorporated in the second edition of the *Principia*. Ramsey was one of the first to expound the early teachings of Ludwig Wittgenstein, by whom he was greatly influenced. In his last papers he was moving toward a modified and sophisticated pragmatism.

## THE FOUNDATIONS OF MATHEMATICS

A stumbling block in the reduction of mathematics to logic attempted in *Principia Mathematica* has long been its appeal to the so-called ramified theory of types, introduced in order to cope with the paradoxes discovered by Russell and others. The excessive restrictions demanded by the theory of types were mitigated by introducing an ad hoc axiom of reducibility, which Ramsey, following Wittgenstein, held to be at best contingently true. Ramsey was one of the first to argue, following Giuseppe Peano, that many of the notorious paradoxes depended on the use of equivocal semantic notions having no place in mathematics. By introducing the notion of “predicative functions”—roughly speaking, truth-functions permitting infinitely many arguments—Ramsey was able to show that the paradoxes could be avoided without appeal to an axiom of reducibility. In order to improve what he regarded as an unsatisfactory conception of identity in *Principia Mathematica*, Ramsey proposed the wider concept of “propositional functions in extension,” considered as correlations, not necessarily definable, between individuals and associated propositions. Fully elaborated, this view would seem to lead to a markedly nonconstructivistic set theory, which most contemporaries would find unacceptable. Ramsey’s distinction between semantic and logical paradoxes and his rejection of that part of the theory of types that subdivides types into “orders” has been almost universally accepted by his successors.

## PHILOSOPHY OF SCIENCE

In a striking paper, “Theories,” Ramsey developed a novel method for eliminating overt reference to theoretical entities in the formal statement of scientific theory. The method consists of replacing, in the axioms of the formal system expressing the scientific theory in question, every constant designating a theoretical entity with an appropriate variable and then applying universal quantification over the propositional matrices thus obtained. Ramsey

was able to show that the conjunction of the universally quantified statements thus derived from the original axioms would have the same observational consequences as the original axiom system. This technique is of interest to philosophers concerned with the ontological implications or commitments of scientific theory.

## PROBABILITY

Ramsey sketched a theory of probability considered as measuring a degree of “partial belief,” thereby providing a stimulus to what are sometimes called “subjective” or “personalistic” analyses of probability. His most important idea was an operational test for degree of belief. Suppose somebody, *P*, has no preference between the following options: (1) to receive  $m_1$  for certain, and (2) to receive  $m_2$  if  $p$  is true but  $m_3$  if  $p$  is false, where  $p$  is some definite proposition and  $m_1$ ,  $m_2$ , and  $m_3$  are monetary or other suitable measures of utility for *P*. Then *P*’s degree of belief in  $p$  is proposed to be measured by the ratio  $(m_1 - m_3)/(m_2 - m_3)$ —roughly speaking, therefore, by the betting odds that *P* will accept in favor of  $p$ ’s being true, given the relative values to him of the possible outcomes.

## GENERAL PHILOSOPHY

Ramsey’s most suggestive idea in general philosophy was that of treating a general proposition, say of the form “all *A*’s are *B*,” as a “variable hypothetical,” considered not as a truth-function (as it had been in his earlier papers) but rather as a rule for judging that if something is found to be an *A* it will be judged to be a *B*—that is, as a formula for deriving propositions in certain ways rather than as an authentic proposition having truth-value. This idea is connected with Ramsey’s unfortunately fragmentary explorations into the connections between belief, habit, and behavior. Ramsey’s papers on facts, propositions, and universals also have not outlived their usefulness.

*See also* Keynes, John Maynard; Logical Paradoxes; Mathematics, Foundations of; Peano, Giuseppe; Philosophy of Science; Pragmatism; Probability; Russell, Bertrand Arthur William; Scientific Theories; Type Theory; Whitehead, Alfred North; Wittgenstein, Ludwig Josef Johann.

## Bibliography

A collection of Ramsey’s work, including previously unpublished papers, was published posthumously as *The Foundations of Mathematics and Other Logical Essays*, edited by Richard B. Braithwaite (London: K. Paul, Trench, Trubner, 1931). This collection has a preface by G. E. Moore, a useful editor’s introduction, and a complete bibliography.

For the definitions of “predicative functions” and “functions in extension,” see especially pp. 39–42, 52–53; Ramsey’s discussion of theories is mainly on pp. 212–236; the generalized betting definition of degree of belief occurs on p. 179.

For discussions of Ramsey’s work, see Israel Scheffler, *The Anatomy of Inquiry* (New York: Knopf, 1963), pp. 203–222, which contains a critical exposition of Ramsey’s procedure for eliminating theoretical terms; Herbert Gaylord Bohnert, *The Interpretation of Theory* (PhD diss., University of Pennsylvania, 1961), further elaboration of Ramsey’s work on the nature of scientific theory; Leonard J. Savage, *The Foundations of Statistics* (New York: Wiley, 1954), which acknowledges indebtedness to Ramsey’s definition of partial belief; and Gilbert Ryle, “‘If,’ ‘So,’ and ‘Because,’” in *Philosophical Analysis*, edited by Max Black (Ithaca, NY: Cornell University Press, 1950; reprinted, New York, 1963), which is a discussion of hypothetical statements as “inference licenses.”

*Max Black (1967)*

## RAMUS, PETER (1515–1572)

Peter Ramus was a logician, educational reformer, and author of many widely used works on philosophy and letters. He was born Pierre de la Ramée in Cuts (Oise), in northern France, the son of an impoverished descendant of a noble family from Liège. After beginning Latin at Cuts, he went to study at Paris, probably between the ages of eight and twelve, and despite grave financial difficulties received his master of arts degree there at the age of twenty-one. His master’s inaugural thesis, according to one still widely circulated but questionable report, was *Quaecumque ab Aristotele Dicta Essent, Commentitia Esse* (Whatever Aristotle has said is a fabrication; the common translation of *commentitia* as “false” is oversimplified).

In 1543, Ramus (he had adopted Petrus Ramus as the Latin form of his name) published two works growing out of his teaching, *Dialecticae Partitiones* (The structure of dialectic,” also titled *Institutiones Dialecticae* [Training in dialectic]) and *Aristotelicae Animadversiones* (Remarks on Aristotle), which violently attacked Aristotle and the university curriculum as confused and disorganized. The university faculty, led largely by doctors of medicine, secured from Francis I a decree forbidding the sale of these books and prohibiting their author from teaching publicly and from writing on philosophy (which included all academic subjects other than grammar, rhetoric, medicine, law, and theology). Ramus, however, quietly continued to teach and write and in 1545 moved to the Collège de Presles in Paris, where he was joined by his earlier asso-

ciate, Omer Talon (Audomarus Talaeus). Ramus soon became principal and dedicated himself, with great success, to promoting more purposeful and effective teaching. In 1547, Henry II lifted the ban against Ramus and in 1551, he appointed him professor of eloquence and philosophy in the body of professors supported by the king, which was later known as the Collège de France; Ramus became its first dean. Earlier an observant Catholic, he embraced the Protestant reform around 1562, withdrawing to Fontainebleau in 1562–1563 during the religious wars and to Rhenish Germany and Switzerland from 1568 to 1570. He returned, however, and was murdered on the third day of the Massacre of St. Bartholomew. Charles Waddington’s assignment of his murder to an academic opponent, the physician Jacques Charpentier, is repeated in many encyclopedia articles but is without demonstrable foundation.

## WORKS

Ramus’s published works run to some sixty-odd titles, supplemented by thirteen additional works of Talon, his frequent collaborator. The works of the two men appeared mostly between 1543 and 1650, in nearly eight hundred (at present) known editions and adaptations (some eleven hundred if works published in collected editions are separately enumerated). Besides the pivotal writings on dialectic, or logic, and on rhetoric, Ramus’s works include classical editions and commentaries; lectures on physics, metaphysics, and mathematics; textbooks for grammar, arithmetic, algebra, and geometry; miscellaneous orations and open letters; and the posthumously published *Commentariorum de Religione Christiana Libri Quatuor* (1576), a basically Zwinglian theological work, unoriginal and apparently of little influence. Other works, notably Latin translations from the Greek, remained unpublished at his death. Although most of his writing was in academic Latin, he published a few works in French, including a *Gramere* of the French language (1562) in a reformed spelling that was developed from that of Louis Meigret.

## PHILOSOPHY

The striking orderliness of Ramus’s philosophy is superficial and is determined by pedagogical serviceability rather than by insight. His *Dialectica* (French, 1555; Latin, 1556, with subsequent revisions), later called also *Logica*, is the key work in the Ramist canon and appeared in nearly 250 extant editions or adaptations, chiefly Latin. The *Dialectica* grew out of his 1543 works and proposed to supplant the highly complex quantified logic of the



Middle Ages, so objectionable to humanists. Actually, it exaggerates—at times grotesquely—the quantifying drives built up in medieval Scholasticism. Following the *De Inventione Dialectica* of Rudolph Agricola, Ramus reduced all argumentation to one “art of discourse” (*ars disserendi*, a Ciceronian definition common during the Middle Ages), which he called indifferently dialectic or logic. He thus did away with dialectic as a separate art that argues from probabilities and is thereby distinct from a scientific logic, which argues from certainties or necessity.

RHETORIC. By the same token, he also dispensed with rhetoric as a separate argumentative art persuading to action. The Ramist *Rhetorica* (1548), published under Talon’s name but with Ramus’s close collaboration (in some 175 known extant editions or adaptations), reduced rhetoric explicitly to mere “ornamentation,” or the application of tropes and figures, conforming to what had been, in fact, a strong trend in medieval thinking about rhetoric. Like Agricola, Ramus treated logic or dialectic as made up of *inventio* (discovery of arguments for any kind of discourse, from mathematics to poetry) and *iudicium* or *dispositio* (the arrangement of arguments, including for Ramus not only syllogism but also method, likewise referable to any and all discourse). Ramus’s treatment of syllogism varied somewhat from some previous treatments but in no original or insightful way, and he did nothing to advance formal logic. Still, his influence was vast and symptomatic.

LOGIC. In the wake of Scholasticism, logic had a high prestige value even among humanists. Ramus made it accessible to all by withdrawing it, more than even medieval Scholasticism had done, from the scientifically elusive world of sound and word and by associating it more with the sense of vision through overt or covert resort to spatial constructs or models in his teaching. Most notable among these models were the dichotomized divisions, often arranged in bracketed tabular form, for analysis of everything under the sun. One divided a subject into two parts, subdivided each of these into two, then again dichotomized each subdivision, and so on. The resulting structure somehow corresponded both to extramental actuality and to the contents of the mind. The intensified passion for this far-from-new procedure was associated with the new medium of typography, which reproduced these and other spatial constructs with an ease and conviction unknown in a manuscript-oriented civilization.

METHOD. In this climate Ramists gave the term *logical analysis* its first extensive currency and developed concern with method. Between 1543 and 1547 the treatment of method earlier found largely in rhetoric manuals had been transplanted into logic manuals published separately by Johannes Sturm and Philipp Melanchthon. During this period Ramus effected the same transplantation in a pseudonymous 1546 revision of his *Dialecticae Partitiones*, from which method made its way into the *Dialectica* from 1555 on. For Ramus, method prescribed treating any subject by going from the general to the particular, although for special reasons one could use cryptic method, proceeding from the particular to the general. Dichotomization implemented method.

Metaphysics was absorbed or displaced by logic, which Ramus passionately but unconvincingly identified with Plato’s dialectic. Ethics was to be taught by methodized analysis of biography and history, and the physics that had formed so great a part of Scholastic philosophy was replaced, in principle at least, by analytic study of works on natural history such as Vergil’s *Georgics*.

## INFLUENCE

Ramus’s realignments involved him in disputes with Antonio de Gouveia, Joachim de Perion, Pierre Galland, Jacques Charpentier, Adrien Turnèbe, Jean Riolan the elder, and Jakob Schegk, disputes protracted after Ramus’s death by hundreds of litigants. Ramist-inspired agitation over method set the stage for René Descartes (who at La Flèche studied a post-Ramist logic textbook with a section on method) and helped make meaningful the application of the nickname “Methodists” to John Wesley’s followers. The modern encyclopedia owes a good deal of its organization to the Ramist and semi-Ramist tradition as represented by polymath organizers of knowledge such as Johann Heinrich Alsted. Ramus’s followers, numbered by the thousands in the sixteenth and seventeenth centuries, were distributed, in descending abundance, through Germany, the British Isles and their American colonies, France, Switzerland, the Low Countries, and Scandinavia. Anti-Ramists such as Nicolas de Grouchy, Everard Digby, and Francis Bacon and Ramists such as Johann Thomas Freige (Freigius), Gabriel Harvey, and John Milton crossbred to produce various syncretists, such as Bartholomew Keckermann, Andreas Libavius, Alsted, and Robert Sanderson. Ramism and its derivatives were particularly popular in Calvinist “middle” or secondary schools for cultural and psychological rather than directly religious reasons: The Ramist account-book interpretation of knowledge and actuality

appealed strongly to the bourgeois mind. Influence in strictly university circles and on speculative thought was more intermittent or indirect, but extraordinarily pervasive.

*See also* Aristotle; Bacon, Francis; Descartes, René; Logic, History of; Medieval Philosophy; Melanchthon, Philipp; Milton, John.

### **Bibliography**

- Ong, Walter J., S.J. "Ramist Method and the Commercial Mind." *Studies in the Renaissance* 8 (1961): 155–172.
- Ong, Walter J., S.J. *Ramus and Talon Inventory*. Cambridge, MA: Harvard University Press, 1958.
- Ong, Walter J., S.J. *Ramus, Method, and the Decay of Dialogue*. Cambridge, MA: Harvard University Press, 1958. Includes an exhaustive bibliography.
- Risse, Wilhelm. *Die Logik der Neuzeit*. Vol. I. Stuttgart: F. Frommann, 1964.
- Waddington, Charles. *Ramus*. Paris: C. Meyrueis, 1855.

*Walter J. Ong, S.J. (1967)*

## RANDOMNESS

*See Chaos Theory; Probability and Chance*

## RASHDALL, HASTINGS

(1858–1924)

The English theologian, philosopher, and historian Hastings Rashdall was born in London, the son of an evangelical clergyman. He was educated at Harrow and at New College, Oxford, where he read Classical Moderations and "Greats." He remained at Oxford two years after graduation, reading philosophy and theology and working on an essay on the history of medieval universities, for which he won the chancellor's prize in 1883. Much of his next twelve years was taken up with expanding this essay for publication in 1895 as a work in three volumes.

In 1883 he left Oxford to become a lecturer at St. David's College, a college for the education of the clergy in Lampeter, Wales, and in December of that year he was appointed a tutor in theology at University College, Durham. In 1889 he returned to Oxford as a fellow of Hertford College and in 1894 was appointed for a year as chaplain and divinity tutor at Balliol, without relinquishing his Hertford fellowship. He returned in 1895 to New College as fellow and tutor and dean of divinity. He retained his New College fellowship but not his tutorship

on his appointment in 1910 as a canon of Hereford Cathedral. He remained in Hereford until 1917, when he became dean of Carlisle, an office he retained until his death.

Rashdall was primarily a theologian and secondarily a philosopher, although he would have been unwilling to draw a clear distinction between the two. His aim was to keep philosophy religious and religion philosophical. Even his history of medieval universities aimed at establishing the rational foundations of religion and ethics, the close connection between the intellectual and spiritual life, and the place of mind in the constitution of the world.

Rashdall justly described himself as "on the left wing of the Church and the right wing of the philosophers." His liberalism in religion and forthright opposition to bigotry kept getting him into trouble with the defenders of orthodoxy. The last years of his life were clouded by the false charge that he denied the divinity of Christ—a charge based on a newspaper misrepresentation of his observation that Jesus never claimed divinity for himself.

Philosophically Rashdall was a personal idealist. Although he held that there is no matter apart from mind—a personal Mind, "in which and for which everything that is not mind has its being"—he rejected monism. Minds are substantial, and every consciousness is exclusive of every other. Individual minds are produced by the eternal Mind, which is God, but are neither included in it nor adjectives of it. In line both with this metaphysical position and with his general distrust of mysticism, Rashdall held our knowledge of God to be inferential.

Rashdall's most important philosophical work is his two-volume *The Theory of Good and Evil*. Although it made no distinctively original contribution to ethics, it is perhaps the best general introduction to the subject written from an objectivist point of view, before the advent of metaethics and the application of philosophical analysis. Rashdall's treatment is thorough and comprehensive, and the book leaves no doubt about the importance for theory and practice of the issues discussed. Although it is not a history of ethics, it includes illuminating expositions and criticisms of theories of classical moral philosophers where these are relevant to the development of his own theme.

Rashdall's emphasis on the value of human personality found expression in his moral theory. Intuitionism, in the sense of acceptance of impersonal moral laws binding independently of their consequences, was wholly

alien to his thought. He was an uncompromising utilitarian, for whom actions are to be judged by their tendency to produce the greatest good or well-being for human beings. There are, indeed, moral intuitions, but they are about the relative value of ends, not about the rightness of rules of conduct. The good that it is the duty of each to produce for all is a personal good but is not confined to pleasure or happiness. Pleasure is only one element that, in interrelation with other mutually modifying elements, including morality, contributes to form an ideally good pattern of life. It was Rashdall who coined the term *ideal utilitarianism* to distinguish this form of the theory from the traditional hedonistic utilitarianism it has generally replaced, partly through his own influence. One advantage of the abandonment of hedonism claimed by Rashdall is that it enables the utilitarian to include in moral judgment the quality of the act itself as well as of its consequences. Thus, the disposition to promote the general good can be taken as itself part of the good to be promoted.

Much of the second volume of *The Theory of Good and Evil* deals with the metaphysical and theological presuppositions of an absolute objective morality. Rashdall held that only in metaphysics can we find an ultimate defense of the validity of moral judgments and that personal idealism has the best chance of supplying it. One postulate of morality is the existence of individual selves to which actions may be attributed; another is the existence of God, as possessing and willing the absolute moral ideal; and a third is immortality. Although he was a determinist, Rashdall escaped having to hold God responsible for evil in human willing because he regarded God not as strictly omnipotent but as limited by those eternal necessities that are part of his own nature.

**See also** Analysis, Philosophical; Idealism; Metaethics; Religion and Morality; Utilitarianism.

### **Bibliography**

Rashdall's works include *The Universities of Europe in the Middle Ages*, 3 vols. (Oxford: Clarendon Press, 1895); "Personality Human and Divine," in *Personal Idealism*, edited by Henry Sturt (New York: Macmillan, 1902), *The Theory of Good and Evil*, 2 vols. (London, 1907; 2nd ed., London: Oxford University Press, H. Milford, 1924); *Philosophy and Religion* (London: Duckworth, 1909); *The Problem of Evil* (Manchester, U.K., 1912); *Is Conscience an Emotion?* (Boston: Houghton Mifflin, 1914); and *The Moral Argument for Personal Immortality* (London, 1920).

For a discussion of Rashdall, see P. E. Matheson, *The Life of Hastings Rashdall* (London: Oxford University Press, H. Milford, 1928).

A. K. Stout (1967)

## RATIONAL INTUITION

See *Intuition*

## RATIONALISM

The term *rationalism* (from the Latin *ratio*, "reason") has been used to refer to several different outlooks and movements of ideas. By far the most important of these is the philosophical outlook or program that stresses the power of a priori reason to grasp substantial truths about the world and correspondingly tends to regard natural science as a basically a priori enterprise. Although philosophies that fall under this general description have appeared at various times, the spirit of rationalism in this sense is particularly associated with certain philosophers of the seventeenth and early eighteenth centuries, the most important being René Descartes, Benedict de Spinoza, and Gottfried Wilhelm Leibniz. It is rationalism of this type that will be the subject of this entry.

Two other applications of the term should, however, be distinguished.

### RATIONALISM IN THE ENLIGHTENMENT

The term *rationalism* is often loosely used to describe an outlook allegedly characteristic of some eighteenth-century thinkers of the Enlightenment, particularly in France, who held an optimistic view of the power of scientific inquiry and of education to increase the happiness of humankind and to provide the foundations of a free but harmonious social order. In this connection "rationalistic" is often used as a term of criticism, to suggest a naive or superficial view of human nature that overestimates the influence of benevolence and of utilitarian calculation and underestimates both the force of destructive impulses in motivation and the importance of such non-rational factors as tradition and faith in the human economy. Jean d'Alembert, Voltaire, and the Marquis de Condorcet, among others, are often cited in this connection. Although there is some truth in these criticisms, the naïveté of these and other Enlightenment writers has often been grossly exaggerated. Also, insofar as "reason" is

contrasted with “feeling” or “sentiment,” it is somewhat misleading to describe the Enlightenment writers as rationalistic, for many of them (Denis Diderot, for example) characteristically emphasized the role of sentiment. Reason was praised in contrast with faith, traditional authority, fanaticism, and superstition. It chiefly represented, therefore, an opposition to traditional Christianity.

Here there are two contrasts with the seventeenth-century rationalism of Descartes and others. First, this rationalism is not characteristically antireligious or non-religious; on the contrary, God in some sense, often in a traditional sense, plays a large role in rationalist systems (although Spinoza’s notion of God was extremely unorthodox, and it is notable that the opposition of reason and faith is important in his *Tractatus Theologico-Politicus*). Second, the view of science held by such Enlightenment thinkers as Voltaire was different from that of rationalism, being much more empiricist. The central contrast embodied in the term *rationalism* as applied to the earlier systems is that of reason versus experience, a contrast that is certainly not present in the Enlightenment praise of the “rational.” Parallel to this difference, there is a difference between the characteristic seventeenth- and eighteenth-century views on the nature and importance of system; the eighteenth century declared itself against the *esprit de système* of the seventeenth century, with its elaborate metaphysical systems, and in favor of an *esprit systématique*, which could be orderly without being speculatively ambitious. (See d’Alembert, *Preliminary Discourse* to the *Encyclopédie* and Condillac, *Traité des systèmes*.)

### RATIONALISM IN THEOLOGY

The Enlightenment spirit of rational criticism directed against the supposed revealed truth of the Scriptures also had effects within Christianity itself. In this connection the term *rationalism* is used in a specific theological sense to refer to the doctrines of a school of German theologians that was prominent roughly between 1740 and 1840, and which had great influence on the development of biblical criticism. With their spirit of antismaterialism can be associated Immanuel Kant’s *Die Religion innerhalb der Grenzen der blossen Vernunft* (1793), in which rational morality is the basis of religious belief.

However, the best-known use of “rationalism” in a religious connection is an entirely negative one, in which it stands for an antireligious and anticlerical movement of generally utilitarian outlook, laying great weight on historical and scientific arguments against theism. This

use of the term, a popular rather than a technical one, seems now to be obsolescent, its place being taken by *humanism*.

### RATIONALISM VERSUS EMPIRICISM

Rationalism as it will be discussed here is standardly contrasted with empiricism. This contrast (which rests on that contrast between reason and experience which has already been mentioned) is now so basic to the use of the terms that no account can afford to ignore it, and a number of comparisons between views associated with these two outlooks will be made in the course of this entry. It is of course impossible to give a detailed comparison of the two outlooks, and in general comparisons will be introduced incidentally to the account of rationalist ideas. There is, however, one issue, that of innate ideas, which embodies a central disagreement between the two, and regarding which an account merely from the rationalist side would be particularly unilluminating. On this issue the disagreements will be considered in rather greater detail than elsewhere. At the same time, it is hoped that the treatment of this issue will give slightly more insight into the rationalist outlook than can be achieved by what is at other points inevitably a very selective summary of rationalist opinions.

### INNATE IDEAS

Descartes distinguished three classes of “ideas” (by which he meant merely whatever it is in a man’s mind in virtue of which he can be said to be thinking of a given thing): adventitious, factitious, and innate. The first type came to the mind from experience, the second were constructed by the mind’s own activity, and the third were created by God together with the mind or soul itself. The last included what were for Descartes the three fundamental ideas of the basic types of substance: God, mind, and matter (or extension). For the most part Descartes argued negatively for the view that these ideas are innate, trying to show that they could not be derived from experience (where this means, fundamentally, sensation).

His argument had two main points. First, the ideas are pure, containing no sensory material; these ideas are not images, reproductions, or copies of sensory experience. Descartes regarded this as fairly obvious in the cases of God and of mind; and he made a particular effort (as in the argument of the wax in *Meditations* II) to establish the same claim for matter. Second, the fundamental ideas implicitly contain, in different ways, some idea of infinity, and in grasping the idea one thereby grasps the possibility of infinitely many and various modifications to which

mind and matter can be subject. In the case of God this argument goes further, for here we grasp an actual infinity of perfections implicit in the idea. The same point, however, holds for all these ideas: The grasp of infinitely many possibilities must transcend what has been given to us in experience, since experience could have given us at best only a limited set of such conceptions, corresponding to what had actually been experienced.

Even if both points of the argument are granted as showing that these ideas are not totally derived from experience, it might be doubted whether they are enough to show that the ideas are innate. For might they not be grasped in some nonempirical manner at a later stage of life—for example, when (or if) someone comes to think in these very general terms? In Descartes's philosophy there is at least an implicit answer to this objection. Descartes thought that the pure ideas of mind and matter are used in the comprehension of experience even before they become conscious in reflection. It is by reference to these ideas that one forms the ordinary unreflective conceptions of oneself as having a series of thoughts or of a material object as enduring, occupying space, and having various characteristics, even though, before reflection, one's conceptions of these things will be confused. Thus, the operation of the pure idea is implicit in ordinary pre-reflective experience, and such experience begins to be acquired from the moment of birth; therefore, there is ground for calling the pure ideas innate.

In the case of God the argument is slightly different, since it is less clear that this idea is "put to use" in any pre-reflective way. Here Descartes may have meant to claim merely that it would be natural to the power and economy of God's operations that he should implant the idea of himself in the soul at its creation, "the mark," as Descartes put it, "of the workman on his work." There is indeed a difficulty in seeing how, for Descartes, there could be an idea in the mind of which the mind is not fully conscious (as this account implies), since for Descartes "mind" and "consciousness" were virtually equivalent. And this difficulty also arises for the ideas of mind and matter, since Descartes explicitly denied (presumably there was no alternative) that the infant or young person is fully conscious of his innate ideas; they are latent and emerge only later—in the process of learning language, for instance. Nevertheless, Descartes's claims for the operation of fundamental ideas in pre-reflective consciousness, although not quite consistent with his metaphysics of the mind, became an important element in later theories of innate ideas, especially in the debate with empiricism.

**INNATE PRINCIPLES.** Descartes appealed only to innate ideas, or concepts, the materials of judgments and beliefs. He did not invoke innate principles, or propositions, his view apparently being that granted innate ideas, we have only to grant in addition a certain power of the mind to elicit features implicit in these ideas in order to explain how necessary knowledge could be derived from innate ideas (as he supposed it could).

Leibniz, however, who continued the Cartesian insistence on innate ideas, added a requirement for innate principles. His argument was of the same general type as that ascribed to Descartes with respect to the ideas of mind and matter: If there were no innate and unlearned propositions, we could learn no propositions at all—at least not by way of logical deduction. For, he argued, confronted with any valid inference of the form " $P$ , so  $Q$ ," we could not see that  $Q$  followed from  $P$  except by having already grasped the necessary truth of the proposition "if  $P$  then  $Q$ ." Thus, in order to follow any inference and to learn anything by deduction, first premises are required that must themselves be unlearned.

An objection to this argument can be seen from the famous difficulty raised by Charles L. Dodgson (Lewis Carroll) that if there is necessarily a difficulty in seeing the validity of the original inference as it stands, the same difficulty will recur with the inference obtained by the addition of the "innate" major premise; to grasp the validity of this inference, another major premise would seem to be required, and so on, thus starting a vicious regress. Dodgson's point makes it clear that no multiplication of premises can be adequate to extricate the validity of an inference; what is needed is something of a different category, a rule. At this point a characteristic empiricist rejoinder to Leibniz's puzzle is to claim that the rules of inference are not unlearned but are learned in the course of learning a language (they are the rules implicit in the correct use of "if," "then," "not," and so on). This illustrates the natural and perhaps inevitable tendency of empiricism, in contrast with rationalism, to turn to a linguistic account of logical necessity. (Such an account, however, even if adequate in itself, may not dispose of the issues as thoroughly as empiricism has tended to believe; the question remains of what is involved in learning a language.)

Leibniz, in introducing the argument just considered, explicitly stated that he was of the "Platonic" opinion that a priori knowledge (at least) is innate and "recollected" (*New Essays*, Book I). There is a difficulty, however, in knowing how far the doctrine is supposed to range: Leibniz's doctrine that the soul is a monad and

that every monad only develops its own inner potentialities, being unaffected by anything outside, implies that in one sense all thoughts, of whatever kind, are innate. This problem involves major questions in the interpretation of Leibniz—in particular, of his views on sense perception. However, it seems reasonable to say that at least in his remarks on innateness in the *New Essays* Leibniz was distinguishing between kinds of knowledge and ideas, such that some (the pure and a priori) can be said to be innate and others cannot.

Leibniz's remarks were in criticism of the First Book of John Locke's *Essay concerning Human Understanding*, and they constitute a subtle consideration of the issues lying between rationalism and empiricism at this point. Locke's First Book, although called "Of Innate Ideas," is in fact chiefly concerned with innate principles (and in some part with the alleged innate principles of morality that had been advanced by his adversary, Lord Herbert of Cherbury). Locke considered various characteristics supposed to show that a given proposition is innate (that it is universally believed, that it is assented to as soon as understood, and so forth), and had little difficulty in showing that these are inadequate. He then turned to the consideration that tiny children do not display elaborate conceptions of logic and mathematics such as are alleged by rationalists to be innate. His principle in this instance was "There is nothing in the mind of which the mind is not conscious"; if these conceptions were innate, they would be in the infant's mind, hence it would be conscious of them and (presumably) could display this consciousness. Leibniz, in reply, claimed that this so patently follows that Locke, in insisting on the principle about consciousness, was in effect begging the question: This principle is what the issue turns on. But, as has been seen, this was not how Descartes put the matter. Leibniz here made the cardinal point of the discussion his own non-Cartesian doctrine of subconscious perceptions (connected with his general doctrine of continuity).

**DEBATE WITH EMPIRICISM.** Once the obvious fact is granted that the allegedly innate ideas do not manifest themselves temporally before other experience, it may be wondered whether any point remains to calling them innate. It has sometimes been suggested that the doctrine of innate ideas merely depends on a confusion between a logical and a temporal sense of "prior." However, this is to underestimate the force of the rationalist claims that the allegedly innate material is such that its operation is a precondition of our learning anything else. It is not easy to decide how to evaluate these claims, as against the central empiricist claim that no such preexisting material

need be postulated (the so-called *tabula rasa* theory of the mind). For one thing, empiricism in its first developments tended to make up for the lack of original raw material by crediting the mind with a very elaborate set of operations. This was evidently the case with Locke, who used such notions as "abstraction," "reflection," and "intuition"; who spoke of "ideas" that are not evidently mere copies from sense perception; and who admitted a non-empirical notion of "substance" and its powers. His position retained a number of rationalist elements. The much more economical apparatus of David Hume, which in effect admits nothing but sensations, their copies, and the operations of association, defines a quite distinctive empiricist theory.

If the debate about innate ideas is cast in terms of a Humean empiricism, there remain principally two issues, one logical and one psychological. The logical issue concerns the question whether highly general concepts, such as those used in mathematics and the sciences, are reducible to or analyzable into those sorts of empirical concepts that can plausibly be said to be derived from sense experience. It would be widely agreed that the answer to this question would be "no." The second, psychological issue is whether the acquisition of concepts, such as occurs in language learning—and this would include even the supposedly straightforward empirical concepts—can be adequately explained by a psychological model postulating only the minimum empiricist requirements of sense perception, retention, association, and so forth. There is influential opinion (held by Noam Chomsky and others) that the answer to this, too, must be "no"; any adequate model may well require stringent innate constraints on the direction and nature of generalization from learning situations. How far these restraints might be supposed to approximate to the rationalists' conceptions of innate ideas—or, in other words, whether the model demands an innate analogue to the possession of concepts—remains to be seen.

If this is indeed an open question, then there is an explicitly psychological version of the rationalist view that is still worth serious consideration. This is not, of course, to say that the innate elements in an adequate model would be likely to correspond to the particular sorts of "ideas" that the rationalists selected for this status—such as the metaphysical notions of God, matter, and mind. Also, there was certainly an endemic confusion, in both the rationalist and the empiricist position on this issue, between psychological and logical issues. Nevertheless, there is still some life in the question, in both its logical and its psychological aspects, the occur-

rence of the psychological term *innate* in the original debate not being merely the result of confusion.

## KNOWLEDGE

It was remarked above that there would now be wide agreement that many general theoretical concepts of mathematics and the sciences do not admit of total reduction to empirical concepts. In contrast with positivist or operationalist views it would be agreed by many that such concepts as “mass,” for instance, are not a mere shorthand for sets of possible observation data. Such agreement, however, although it would constitute a rejection of strict empiricism, would not in itself constitute an acceptance of rationalist views about such concepts. It is possible to think that these concepts “transcend,” or “go beyond,” the empirical merely in virtue of conventional elements—that they are parts of humanly constructed models of reality which relate the observable by imposing a structure on it.

Essential to rationalism, however, is a realistic view (incompatible with even a modified empiricism) about the relation of these concepts to reality and about the necessary relations obtaining between these concepts themselves. The intellectual grasp of these concepts and the truths involved in them is seen as an insight into an existing and unique structure of the world. It is not easy to express this picture (which in varying degrees dominated the rationalists) in less figurative language, but the picture has at least two consequences: that there is a unique set of concepts and a unique set of propositions employing these concepts that adequately express the nature of the world, and that these propositions form a system and could ideally be recognized as a set of necessary truths. There are, admittedly, difficulties about the last point, particularly with reference to Leibniz (these will be considered in the next section). However, something like this general picture is central to rationalism and leads immediately to the question of how anyone can come to know this uniquely correct representation of the world. This invites two more specific questions: What, in general, is the guarantee that knowledge of the world is possible? how can any individual tell in a particular case whether he has hit on some genuine piece of knowledge?

**DESCARTES'S EPISTEMOLOGY.** Most rationalists tended to answer the first of the above questions by referring to God; some, but not all, did the same for the second; and they varied in the priority that they assigned to the two questions. Descartes started famously with the second question and found the answer in the “clear and distinct

perceptions” of the intellect. Proving, as he supposed, the existence of God via clear and distinct perception, he then employed God’s perfection of “being no deceiver” to establish in general terms the reliability of beliefs that went beyond clear and distinct perception. He was, however, so impressed by the thought that it was only in virtue of humanity being created and sustained by God that he could know anything at all, that he was constantly tempted to double back and use the divine perfection to guarantee even the basic clear and distinct perceptions, thus laying himself open to the charge of arguing in a circle.

However this may be, it is notable that in Descartes “clear and distinct perception” is a thoroughly epistemological category. The truths that can be clearly and distinctly perceived do not constitute one homogeneous logical or metaphysical class of truths; the class includes at least statements of contingent existence (his own, in the *cogito*) and of necessary existence (that of God), contingent statements about immediate psychological experience, and necessary truths about the relations of ideas. The status of these last, which Descartes called eternal truths, is somewhat obscure. Descartes held, in the Augustinian-Scotist tradition, that they were the products of God’s will; but it is left unclear what it is that God has brought about in creating eternal truths, and hence what it is that one knows in knowing them.

**THE CARTESIAN TRADITION.** The development of the Cartesian tradition within rationalism tended to emphasize to an even greater extent the theological elements in Descartes’s theory of knowledge. Thus Nicolas Malebranche retained for the individual case the test of “clear and distinct perception” in a style that seems to assimilate it to moral perception and the promptings of conscience: “One should never give one’s complete assent except to propositions which seem so evidently true that one could not reject them without feeling an interior pain, and secret reproaches of the reason” (*De la recherche de la vérité*, I, Ch. 2; for the moral analogue, see Bossuet, *Traité de la connaissance de Dieu et de soi-même*, Ch. 1, Sec. 7).

Malebranche gives a strongly Augustinian and indeed Neoplatonist turn to the general account of God’s guarantee of the possibility of knowledge. His doctrine was that all our knowledge of the external world is mediated by God; the mind of God contains paradigm ideas in whose form he created the world, and it is these same ideas of which we are conscious when thinking about the world. This is the meaning of Malebranche’s saying that we see all things in God. This doctrine, apart from serving religious purposes, was also an attempt to get around

the difficulties inherent in Descartes's own causal account of relations between matter and mind (which will be considered more generally later in this entry).

The role of God in the foundations of knowledge takes different and less extreme forms in other areas of the rationalist tradition. The greatest contrast to the Malebranche development of Cartesianism might plausibly be said to be Spinoza's system. It is true that Spinoza did assert that it is the nature of God that guarantees the correspondence of our thoughts to the world, but he so transmuted the notion of God that the doctrine is only verbally similar to Cartesianism. "God" is one name ("Nature" is another) for the one substance, that is, everything that there is. This substance has infinitely many attributes, of which we can comprehend only two, mind and matter. These two attributes are necessarily parallel to one another, and corresponding to any mode of the one attribute there must be a mode of the other. Hence, thought and the material world are inherently adjusted to one another, and the development of knowledge consists in the project of rendering the thought component of this relation as clear (in Spinoza's term, as "active") as possible. It admittedly remains obscure how, within the constraints of Spinoza's determinism, this can be regarded as a "project" at all. Despite this and the other notorious difficulties, Spinoza's system is particularly interesting in the present connection as a thoroughgoing attempt to answer the crucial question that was left very much in the air in Descartes's thought, namely, how any knowledge of a necessary truth, regarded as knowledge of the relations of ideas, could also constitute knowledge of the world.

Leibniz's system, for all its radical differences from Spinoza's, resembles it in one respect having to do with the foundation of knowledge: The general possibility of the correspondence of thought to the world is guaranteed metaphysically by the existence of a correlation between the two. The monads are not affected by anything outside and each develops its own activity from within, but a correspondence between the activities of the monads is given by the "Prestablished Harmony"; and knowledge, the correspondence between "conscious" states of certain monads and other monads, is a special case of this. The Prestablished Harmony, however, depends on God's optimal choice, that is, on God's benevolence. Thus, in a less explicitly epistemological form, Leibniz (in contrast with Spinoza) reverted to the original Cartesian standpoint, in that there is a transcendent and personal God who has a will, and it is a result of his will that there is an ultimate guarantee of the possibility of knowledge.

In general, however, Leibniz was not much concerned with epistemological problems; in particular, he was uninterested in the question that was the starting point for Descartes: How can the individual be certain of the truth of anything? Spinoza was concerned with this question, and tried to develop a theory of knowledge that would avoid the regress latent in Descartes's method, arising from the question of how one knows that one knows. In Spinoza's "degrees of knowledge" it is an essential property of the highest, or intuitive, degree that it is self-guaranteeing. Even so, there is an evident shift in the Spinozistic outlook away from the Cartesian question "What do I know, and how do I know it?" Spinoza, like Leibniz and many other rationalists, gave the metaphysical description of the world from "outside," from a "God's-eye" standpoint rather than from the subjective epistemological standpoint from which Descartes (although unsuccessfully) tried to work. It is, perhaps, a mild irony of the history of philosophy that Descartes's attempt to start with subjective questions of epistemology and to "work out" from there had more influence on the development of empiricism than on later rationalism.

## SCIENCE AND SCIENTIFIC METHOD

No attempt will be made here to give an account of the detailed developments of the philosophy of science within rationalist thought, or of the actual scientific conceptions held by or associated with rationalists, although these are of course of great importance, most notably in Leibniz's critique of Cartesian physics and in the development of his concept of force. We shall consider only one or two general points about the rationalists' conception of a completed science and associated notions of scientific method.

Rationalist developments in these matters can usefully be seen in the light of an unresolved conflict within Descartes's system between the method of approaching scientific inquiry and the expected shape of the final product. Descartes favored in principle an approach to inquiry that might be called systematically exploratory. This he called the analytic method; and the straightforward exposition of the results of such an inquiry would be heuristic in style, explaining the resolution of difficulties as they were encountered in the systematic progress. He seems, however, also to have had a picture of a completed science as a complete deductive system, ideally expressed in a unique system of theorems with necessary truths (of a metaphysical character) as its axioms; this he termed the synthetic method of exposition. There is, perhaps, no essential clash between these two ideas of



method and result; but Descartes seems not to have been clear about the relation between the two or how this specific method, fully pursued, would yield this specific result. Ambiguities about this question emerge in Descartes's accounts of the role of experiment, in which he sometimes gives the incoherent impression that he is both engaged in logical deduction of scientific laws from self-evident metaphysical premises and doing experiments to assist him in this deduction. On the whole, it is probably better to regard the idea of a complete formally deductive metaphysico-scientific system as less important in Descartes's thought than is sometimes supposed, and to see him as using certain limiting principles of scientific explanation, within which he constructs models to explain particular phenomena.

The idea of the total deductive system, however, had a powerful effect on rationalism and reached its most extreme expression in the work of Spinoza, where the "synthetic" method of Euclidean demonstration is explicitly regarded as necessary to the highest form of understanding. This was not just an expository preference; it was an expression of the basic Spinozistic outlook, which regarded the relation of cause to effect as that of logical ground to consequence—for Spinoza all explanatory relations were logical and timeless. The parallel orders of thought and matter, remarked on earlier, supposedly guarantee that the logical relations of ideas will constitute a totally adequate expression of the nature of the world. (A singular application of this notion of total parallelism is to be found in Ehrenfried Walter von Tschirnhaus, who in *Medicina Mentis* [1687] argued that an adequate definition of laughter should be able to produce laughter.)

Leibniz, partly under the influence of Erhard Weigel, was also attracted to the "geometrical method." He devoted a good deal of effort to the project of a universal calculus, which would enable arguments on any subject matter to be cast into a rigorous demonstrative form. However, the idea of such a calculus in no way presupposes an ideal of being able to demonstrate scientific truths from metaphysical or other supposedly self-evident axioms, which was the Spinozistic and, on occasion, the Cartesian ideal. Even if Leibniz started with the notion that it should be possible to settle any argument by appeal to the self-evident, he abandoned it in his mature philosophy, in which he made fundamental the distinction between "truths of reason," which can be established by logical insight on the basis of the law of noncontradiction, and "truths of fact," which depend on the principle of sufficient reason and cannot be established on logical grounds alone. There are some notori-

ous difficulties about this distinction, especially concerning the question of the nature of the contingency of "truths of fact," since Leibniz also held the further general principle that in all true propositions the predicate is contained in the subject. It does seem clear, however, that there is an ineliminable contingency about "truths of fact," and hence that the aspiration of reducing all knowledge to a system of deductions from self-evident premises must be impossible in the Leibnizian system.

Francis Bacon said in his *Cogitata et Visa* (1607), "Empiricists are like ants, they collect and put to use; but rationalists, like spiders, spin threads out of themselves." Bacon, of course, preferred the ants. Although there is some element of truth in the image of the spider, as applied to some rationalist thinkers, it does less than justice to the substantial empirical work done under rationalist inspiration. This is all the more so if one counts Galileo Galilei's view of science as fundamentally rationalist. He certainly rejected any kind of Baconian empiricism and shared the rationalist vision of a mathematical structure of reality that intellectual insight could grasp; but he perhaps had a more sophisticated feeling than any of the philosophers for the balance of imagination and experiment in physics. The rationalist tradition certainly embodied fundamental insights (lacking in empiricism) about the nature of science; above all, it saw the importance of mathematical structures in physical explanation and the vital possibility of a theory's making a conceptual jump beyond the observations and not merely (as in empiricism) an advance in generality. Its sense of the activity of the scientific mind, of its restructuring of observations through concepts and models, was very significant. At the same time, empiricism rightly fought for a clearer distinction between pure mathematics and natural science, undermined the aspirations to final certainty that dogged the rationalists, and emphasized the role of laborious observation and experiment in contrast with the rather dreamlike quality of rationalist visions of the universe. No clearer case exists in the history of philosophy of the need for, and eventual occurrence of, a synthesis; one aspect of that synthesis is neatly summed up in a remark of Giorgio de Santillana that "the true scientist has an empiricist conscience and a rationalist imagination."

## SUBSTANCE AND CAUSALITY

In the history of classical empiricism the concepts of substance and of active causal power together became progressively weaker and were finally abandoned. Thus Locke employed the full Cartesian array of both material

and mental substances, both possessing causal power; George Berkeley banished material substance, partly on the ground that it could not be conceived of as possessing causal activity, which belongs only to mental substance; Hume maintained that the notions of substance and of causal activity are unintelligible. By contrast, in the rationalist tradition the notion of substance has not declined; developments in the idea of causal activity, although partly parallel to the idea of substance, are very different; in general the fortunes of “substance” and of “causal activity” have not been directly linked, as they have proved to be in empiricism—both have undergone considerable and partly independent variations.

In the case of substance (which will be very briefly considered here) the concept has not so much been criticized as used in differing ways to express differing metaphysical views of the world. On one measure, at least, the extremes in this respect are represented by the philosophies of Spinoza and Leibniz. Spinoza gave what he claimed was an a priori demonstration that there could be only one substance (*Deus sive Natura*, God or Nature); this was intrinsically neither material nor mental, these distinctions arising (as noted above) only at the level of the different attributes of this same substance. Essential to Leibniz’s outlook, on the other hand, was an infinite set of substances, the monads, each of them different from all the others. In their character, although there are difficulties of interpretation on this point, they are more of a mental than of a material kind.

On the question of causality an important stream in the history of rationalism stems from the problem left by Descartes, concerning the causal interaction of mind and matter. Descartes’s own view, which postulated simple efficient causation as holding between the two types of substance, failed to appeal to even the most ardent Cartesians, and their attention was particularly directed to this question, although difficulties about the meaning of causation even between material bodies also were considered. The natural tendency in the Cartesian tradition was to move toward attributing all causal power to God, and this movement of thought culminated in the doctrine of occasionalism—that both physical and mental events in the world are occasions for the application of God’s power, which itself directly produces what would normally be called the effects of those events. This doctrine is most thoroughly expressed in the writings of Malebranche. Similar views, however, are to be found in Louis de la Forge (*Le traité de l’esprit et de l’homme*, 1666) and Géraud de Cordemoy (*Le discernement du corps et de l’âme*, 1666), whose work was known to Malebranche.

The theory of occasionalism can be usefully contrasted with Berkeley’s empiricist account of causation. For both the only genuine activity was spiritual. For Berkeley the effects of such activity were also spiritual (mind can affect only mind), and indeed there was no other type of substance. The occasionalists retained material substance and did not find it unintelligible that mind can act upon matter; however, they held that the only mind for which such action is intelligible is the infinite mind of God. Here, as elsewhere, the questions of the gulf between mind and matter and of causation as activity emerge as of common concern to both rationalist and empiricist metaphysics, the influence of Descartes being clearly discernible in both.

Another writer who inclined to occasionalism was Arnold Geulincx (*Ethics*, 1665; 2nd ed., 1675); however, he also suggested a different model for causality, in which God did not, as in occasionalism, make a constant series of miraculous interventions into the natural order but had established ab initio a series of coordinated developments, the relations between which are what is taken for causal interaction. In this connection Geulincx introduced the example of the two clocks, perfectly adjusted to keep the same time, one of which strikes when the other shows the hour; the appearance of causal connection between them is only a result of precise prearrangement.

This same analogy was frequently employed by Leibniz in explaining his own very thoroughgoing version of this thesis, in which all appearance of causal interaction is an instance of the preestablished harmony between the several developments of the monads. Here again there is a notable contrast with and a similarity to empiricism: Both Leibniz and Hume, each representing the culmination of one of the two traditions in its classical form, deny the existence of “transeunt action” between different things and see what is called causation as a correlation between phenomena. Leibniz, however, emphasized some kind of spontaneous activity within the monad, while for Hume neither such activity, nor any notion of a substance, such as a monad, was acceptable. The views of these two philosophers are also worthy of comparison on other subjects, such as space and time; and the points of contact between them are the more significant in the light of the radical and very obvious differences in the spirit, method, and presuppositions of their two philosophies. These differences in the two culminating figures constitute a paradigm, almost a caricature, of the divergent styles of thought associated with rationalism and empiricism, while at the same time similar pressures in the his-

tory of thought produced partly parallel developments in each.

**See also** Alembert, Jean Le Rond d'; A Priori and A Posteriori; Augustinianism; Bacon, Francis; Berkeley, George; Carroll, Lewis; Cartesianism; Chomsky, Noam; Condorcet, Marquis de; Cordemoy, Géraud de; Descartes, René; Diderot, Denis; Empiricism; Encyclopédie; Enlightenment; Experience; Geulincx, Arnold; Herbert of Cherbury; Humanism; Hume, David; Innate Ideas; Kant, Immanuel; Knowledge, A Priori; Leibniz, Gottfried Wilhelm; Locke, John; Malebranche, Nicolas; Neoplatonism; Propositions; Reason; Scientific Method; Scotism; Spinoza, Benedict (Baruch) de; Spinozism; Tschirnhaus, Ehrenfried Walter von; Voltaire, François-Marie Arouet de.

### Bibliography

For an understanding of rationalism, writings by and about particular rationalist philosophers should be read; references to these will be found in the bibliographies of the appropriate entries.

#### GENERAL WORKS

- Bouillier, Francisque. *Histoire de la philosophie cartésienne*. 3rd ed. Paris, 1868.
- Cassirer, Ernst. *The Philosophy of the Enlightenment*. Translated by Fritz Koelln and James Pettegrove. Princeton, NJ: Princeton University Press, 1951. As the title indicates, this is not directly on seventeenth-century rationalism but relates to it.
- Cottingham, John. *Rationalism*. London: Paladin, 1984.
- Doney, Willis. "Rationalism." *Southern Journal of Philosophy*, Supp. 21 (1983): 1–14.
- Lecky, W. E. H. *History of the Rise and Influence of the Spirit of Rationalism in Europe*. London: Longman, Green, Longman, Roberts, and Green, 1865; rev. ed., London, 1910.
- Santillana, Giorgio de, and Edgar Zilsel. "The Development of Rationalism and Empiricism." In *International Encyclopedia of Unified Science*. Vol. II. Chicago: University of Chicago Press, 1941.
- Windelband, Wilhelm. *A History of Philosophy*. Translated by J. H. Tufts. New York, 1901. Part IV, Ch. 2; Part V, Ch. 1. Cites the works and—summarily—the views of many philosophers, including minor ones.

#### PHILOSOPHY OF SCIENCE

- Burt, E. A. *The Metaphysical Foundations of Modern Physical Science*, rev. ed. New York, 1952; paperback reprint, Garden City, NY: Doubleday, 1954.
- Dijksterhuis, E. J. *The Mechanization of the World Picture*. Translated by C. Dikshoorn. Oxford: Clarendon Press, 1961.
- Gillispie, C. C. *The Edge of Objectivity*. Princeton, NJ: Princeton University Press, 1960.

#### INNATE ELEMENTS IN LANGUAGE ACQUISITION

- Fodor, J. A., and J. J. Katz, eds. *The Structure of Language*. Englewood Cliffs, NJ: Prentice-Hall, 1964. Chs. 21 and 22.

Papers by Noam Chomsky and E. H. Lenneberg, respectively.

#### OTHER RECOMMENDED TITLES

- Bealer, George. "Modal Epistemology and the Rationalist Renaissance." In *Conceivability and Possibility*, edited by Tamar Szabo Gendler. Oxford: Oxford University Press, 2002.
- BonJour, Laurence. *In Defense of Pure Reason*. Cambridge, U.K.: Cambridge University Press, 1998.
- BonJour, Laurence. "A Rationalist Manifesto." *Canadian Journal of Philosophy*, Supp. 18 (1992): 53–88.
- Brink, David. "Kantian Rationalism: Inescapability, Authority, and Supremacy." In *Ethics and Practical Reason*, edited by Garrett Cullity. New York: Clarendon Press, 1997.
- Carruthers, P. *Human Knowledge and Human Nature*. Oxford: Oxford University Press, 1992.
- Gendler, Tamar Szabo. "Empiricism, Rationalism, and the Limits of Justification." *Philosophy and Phenomenological Research* 63 (2001): 641–648.
- Markie, Peter. "Rationalism vs. Empiricism." Stanford Encyclopedia of Philosophy. Available from <http://www.plato.stanford.edu>.
- Peacocke, Christopher. *The Realm of Reason*. Oxford: Clarendon Press, 2004.
- Peacocke, Christopher. "Three Principles of Rationalism." *European Journal of Philosophy* 10 (2002): 375–397.

**Bernard Williams (1967)**

*Bibliography updated by Benjamin Fiedor (2005)*

## RATIONALISM IN ETHICS (PRACTICAL-REASON APPROACHES)

Practical-reasoning theory is a kind of metaethical view—alongside noncognitivism and other cognitivism such as naturalism and rational intuitionism—that aims to understand ethics as rooted in practical reason.

Tradition divides the faculty of reason into two parts: theoretical and practical. Theoretical reason concerns what we should believe, practical reason what we should do. Beliefs aim to represent reality and are mistaken or in error when they do not. Theoretical reason's task, therefore, is to discover what is true of the independent order of fact to which belief is answerable. But what about practical reason? What could make it the case that an action is something a person ought to do?

Plainly, ethical convictions also aim at a kind of objectivity. If Jones thinks he should devote all his resources to conspicuous consumption but Smith thinks that Jones should donate some to help the poor, their convictions conflict. Only one, at most, can be true.

Practical-reasoning theories aim to explain the objective purport of ethical conviction, but in a way that respects a fundamental distinction between theoretical and practical reason. Like noncognitivism, these theories sharply distinguish between ethics and those theoretical disciplines that aim to represent some independent reality, whether the order of nature or some supersensible metaphysical realm. They therefore reject both naturalism and rational intuitionism. But they also deny noncognitivism, since they hold that ethical propositions can be true or false.

According to practical-reasoning theories, objectivity consists not in accurate representation of an independent order, but in demands that are universally imposed within an agent's own practical reasoning. What marks ethics off from science is its intrinsically practical character, its hold on us as agents. It is because there is such a thing as practical reason, a form rational agents' deliberations must take, that there is such a thing as ethics.

But what form does rational deliberation take? Uncontroversially, practical reasoning includes reasoning from ends to means. The interesting debates concern what else it involves, if anything, and how instrumental reasoning is itself to be understood. Humeans maintain that means-end reasoning exhausts practical reason and that instrumental reason can be reduced to the use of theoretical reason in discovering means to ends. They tend not to be practical-reasoning theorists, however, since they argue that ethics fundamentally concerns what engages human sympathy or moral sentiment rather than what it is rational for a person to do. By contrast, practical-reasoning theorists deny that practical reason can be reduced to theoretical reason. As Christine Korsgaard has argued, even instrumental practical reason directs an agent who has already used theoretical reason in determining that *B* is the only means to his end *A* to undertake *B* (or to give up *A* as an end). In this way instrumental practical reasoning parallels the structure of *modus ponens* in theoretical reasoning (the move from "*p*" and "if *p*, then *q*" to "*q*").

Pursuing the analogy with theoretical reasoning (while insisting on irreducibility) further suggests that instrumental reasoning cannot exhaust practical reason. When we reason from our beliefs—for example, with *modus ponens*—we reason from their contents, not from the fact of our believing them. We reason from *p* and if *p*, then *q*, not from the facts that we believe that *p* and that we believe that if *p*, then *q*. Similarly, when we adopt an end, we do not simply select it by sheer fiat. Rather, we choose it as something (we think) there is some reason to

do. Thus, when we reason from our ends, we do not reason from the fact that they are our ends but from our commitments to them as things it makes sense to do. That is why instrumental rationality is so uncontroversial. As R. M. Hare argued, it is questionable at best that it follows from the facts that a person's end is to kill someone in the most grisly possible way and that using a cleaver is such a way that the person ought, or has some normative reason, to use a cleaver. What is uncontroversial is simply that the support of reasons transfers from end to means, other things being equal, and from not taking the (only available) means to renouncing the end, other things being equal. It follows only that a person ought to use a cleaver or give up my end.

On grounds such as these, practical-reasoning theorists tend to hold that instrumental rationality cannot exhaust practical reason. But how are we to deliberate about ends? What makes something a reason for adopting an end? Since they hold that reasons for action are necessarily connected to the agent's deliberative perspective, practical-reasoning theorists generally adopt what Korsgaard has called the internalism requirement, according to which a reason must be something the agent could, in principle, be moved by in deliberation and act on. This makes it a necessary condition of something's being a reason for an agent that she would be moved by it insofar as she deliberated rationally.

But what then is rational deliberation? Practical-reasoning theorists are loath to derive a deliberative ideal by independently specifying paradigm reasons for acting and holding that deliberation is rational when it responds appropriately to them. That would theorize practical reason too much on the model of theoretical reason. Rather, they maintain that rational deliberation must be understood formally, so that reason for acting is a status consideration inherit when it is such that it would move an agent who formed her will in accordance with that deliberative ideal.

The aspects that have been considered so far are relatively common among practical-reasoning theories, although not, perhaps, universal. Within these theories, however, there is a major division between neo-Hobbesians and neo-Kantians. Although nothing on the surface of practical-reasoning theory might suggest this result, it is notable that both camps attempt to vindicate the commonsense idea that moral obligations are supremely authoritative. Both argue that (at least some central) moral demands are demands of practical reason.

## NEO-HOBBSIANISM

Recent versions of this view have their roots in ideas advanced by Kurt Baier in the late 1950s and attempt to address a significant problem faced by Baier's early view. Baier argued that reasons for acting must ultimately connect with the agent's interests. This does not reduce all practical reasoning to prudential reasoning, since other forms may advance agents' interests also. Specifically, Baier argued that morality may be viewed as a system of practical reasoning that is in the interest of everyone alike. Since it is mutually advantageous for everyone to regard moral obligations as supremely authoritative, Baier concluded that they actually do create overriding reasons for acting.

David Gauthier objected to Baier's theory that, while it is in the interest of each that all regard interest-trumping moral reasons as supreme, it is unclear how this can show that an individual agent should so regard them, since it will still most advance her interest to act prudentially when morality conflicts with self-interest. Why, then, might it not be true that instrumental and prudential reasoning exhaust practical reason, even if a person should hope to live in a world in which other people view things differently and (mistakenly) treat moral reasons as authoritative?

Gauthier is himself responsible for the major recent neo-Hobbesian practical-reasoning theory. Like Baier, Gauthier begins from the premise that practical reasoning must work to advance the agent's interests, although here his account is more nearly "internalist," since he understands a person's interests to consist in what she would herself prefer were she to be fully informed. Also like Baier, Gauthier argues that the fact that mutual advantage may require individuals to constrain their pursuit of self-interest can be used to show that practical reason counsels this constraint. However, it is not enough that it be true that everyone would do better if everyone so constrained his or her prudential reasoning. The crucial point for Gauthier is that individuals can do better if they constrain self-interest by a willingness to abide by mutually advantageous agreements.

Two agents who appear to each other to be unconstrained pursuers of self-interest simply cannot make agreements, however mutually advantageous the agreements might be, if these agreements would require the agents to act contrary to their own interests. In what have come to be known as prisoner's dilemma situations, therefore, mutually advantageous rational agreement between such persons is impossible. If each believes the other will rationally defect from the agreement on the

condition that doing so is in her interest, then neither can rationally make the agreement.

Personal advantage therefore counsels presenting oneself to others as someone who is not an unconstrained maximizer of self-interest. Of course, it is possible, theoretically, for someone to do this while still deliberating as an unconstrained prudential reasoner. But it may not be practically possible, Gauthier argues, at least not for normal human beings. Human motivation may be sufficiently translucent—through involuntary response, for example—so that the least costly way of appearing to others as someone who can be relied upon to keep mutually advantageous, interest-constraining agreements is actually to be such a person. If that is so, then instrumental and prudential reason will not support themselves as principles to guide rational deliberation. On the contrary, they will recommend that agents deliberate in terms of an alternative conception of practical reason that counsels keeping mutually advantageous agreements, even when this is contrary to self-interest.

As a practical-reasoning theorist, Gauthier believes that reasons for acting cannot be understood except in relation to what should guide a rational agent in deliberation. And he believes that a rational agent is someone whose dispositions of choice and deliberation serve her best and most advance her interest. But just as indirect forms of ethical consequentialism, such as character- and rule-consequentialism, face the objection that they are unstable and threaten to collapse into either act-consequentialism or deontology, Gauthier's indirect consequentialist theory of rationality may face the same objection. What motivates the move away from unconstrained prudence, on the grounds that it cannot support itself in the agent's practical thinking, is a view about the role a principle of rational conduct must be able to play in the deliberations of an autonomous rational agent that may be more Kantian than Hobbesian in inspiration.

## NEO-KANTIANISM

This contemporary tradition may be held to date from Thomas Nagel's *The Possibility of Altruism* (1970) and John Rawls's reinvigoration of Kantian moral and political philosophy in *A Theory of Justice* (1971) and "Kantian Constructivism in Moral Theory" (1980). Nagel's book was read as having both a modest and a more ambitious agenda. His more modest goal, suggested by his title, was to show how such "objective" (or, as he later termed them, "agent-neutral") considerations as "that acting would be relative *someone's* pain" can be genuine reasons for acting. A consideration can be rationally motivating, he argued,

even if the agent lacks any relevant desire for acting on it other than one that is motivated by the awareness of that very consideration. A person may be moved, for instance, by considering long-term interests. And if motivation at a distance is possible with prudence, it can happen with altruism as well. Altruistic and other agent-neutral considerations can be rationally motivating.

Nagel's more ambitious agenda was to argue that practical reasoning is subject to a formal constraint that effectively requires that any genuine reason for acting be agent-neutral. Stressing the "motivational content" of genuine practical judgments, Nagel argued that avoiding a kind of solipsism is possible only if an agent is able to make the same practical judgment of himself from an impersonal standpoint as he does from an egocentric point of view. Since accepting practical judgments from one's own point of view normally motivates, Nagel maintained, making the same judgment of oneself from an impersonal standpoint should normally motivate also. But this will be so only if the reasons for acting that ground practical judgments are agent-neutral. So it is a necessary condition for avoiding practical solipsism that agents take considerations such as that something will advance their own ends or interests as reasons only if they regard them as instantiating more general, agent-neutral reasons, such as that acting will advance someone's ends or interests. Nagel later retreated from this strong claim in a direction that is arguably even more Kantian. Autonomous agency, he later argued, involves an agent's acting on reasons she can endorse from an objective standpoint, and such a set of reasons will include both agent-relative and agent-neutral ones.

Neo-Kantian practical-reasoning theories have been put forward by a number of philosophers, including Alan Gewirth, Stephen Darwall, and Christine Korsgaard. Korsgaard's sympathetic reconstruction of Immanuel Kant's own arguments in a series of papers has been especially influential. Common to all these neo-Kantian approaches has been the idea that the practical reasoning of an autonomous agent has a formal structure, with its own internal standards and constraints, and that these provide the fundamental truth and objectivity conditions for ethical thought and discourse. Thus, Gewirth maintains that fundamental moral principles are derivable from propositions to which a rational agent is committed from within the deliberative standpoint in acting. And Korsgaard argues that even instrumental theorists are committed to the "hypothetical imperative" as a practical norm. Since, however, we regard ourselves to be free as agents to adopt and renounce ends, practical reason can-

not possibly be exhausted by any mere consistency constraint, such as the hypothetical imperative. It follows, the neo-Kantians argue, that practical reason requires norms to regulate the choice of ends no less than to guide the choice of means. In choosing ends for reasons we commit ourselves implicitly to principles of choice as valid for all. But such a commitment is not, they claim, a hypothesis about some independently existing order of normative fact to which we might have cognitive access. That, after all, is precisely the difference between theoretical and practical reason. So the standards to which deliberation is subject must ultimately be based on some formal principle of impartial endorsement that is internal to free practical reasoning itself. And this will be so, they conclude, only if practical reasoning is regulated by some such principle as the categorical imperative, which requires that one act only on principles that one can will to regulate the deliberation and choices of all. If moral demands are ultimately grounded in the categorical imperative also, it will follow that moral demands are demands of practical reason.

**See also** Baier, Kurt; Consequentialism; Decision Theory; Gewirth, Alan; Hare, Richard M.; Intuitionism; Kant, Immanuel; Metaethics; Nagel, Thomas; Naturalism; Noncognitivism; Practical Reason; Rationality; Rawls, John; Reason.

### Bibliography

- Baier, K. *The Moral Point of View*. Ithaca, NY: Cornell University Press, 1958.
- Darwall, S. *Impartial Reason*. Ithaca, NY: Cornell University Press, 1983.
- Falk, W. D. *Ought, Reasons, and Morality*. Ithaca, NY: Cornell University Press, 1986.
- Gauthier, D. "Morality and Advantage." *Philosophical Review* 76 (1967): 460–475.
- Gauthier, D. *Morals by Agreement*. Oxford: Clarendon Press, 1986.
- Gewirth, A. *Reason and Morality*. Chicago: University of Chicago Press, 1978.
- Korsgaard, C. "The Source of Normativity." In *The Tanner Lectures on Human Values*, edited by G. Peterson. Salt Lake City: University of Utah Press, 1994.
- Nagel, T. *The Possibility of Altruism*. Oxford: Clarendon Press, 1970.
- Nagel, T. *The View from Nowhere*. New York: Oxford University Press, 1986.
- Rawls, J. "Kantian Constructivism in Moral Theory." *Journal of Philosophy* 77 (1980): 515–572.
- Rawls, J. *A Theory of Justice*. Cambridge, MA: Belknap Press of Harvard University Press, 1971.

**Stephen Darwall (1996)**

## RATIONALISM IN ETHICS [ADDENDUM]

Moral rationalism, like many philosophical “isms,” is an umbrella term for a variety of constituent claims. Not all moral rationalists endorse all of these claims, but the central ones that have been associated with moral rationalism are:

- 1) The metaphysical thesis: Basic moral requirements are constituted by the deliverances of sound practical reason.
- 2) The epistemological thesis: Humankind’s basic moral requirements are knowable a priori.
- 3) The normative thesis: Moral requirements entail excellent reasons for action.

The metaphysical thesis is at the heart of practical reason approaches to ethics, such as those of Thomas Hobbes and Immanuel Kant, and, in modern times, David Gauthier, Christine Korsgaard, and Michael Smith. Though the views of these thinkers differ in many important respects, they agree that our basic moral duties are a function of sound practical reason. This means that something is a moral duty for a person just because it would be regarded as rationally authoritative were that person (or some idealized counterpart thereof) to reason soundly from his or her most important commitments.

What distinguishes adherents of the metaphysical thesis from one another are their views about the nature of our fundamental commitments, and about what sound practical reasoning looks like. Kantians, such as Korsgaard, will consider some of our fundamental commitments as intrinsic to human nature, and so shared by all human moral agents. Successful reasoning on this shared basis will yield a set of universal moral duties. Hobbesians, such as Gauthier, deny that humans have any essential or intrinsic ends. Yet Gauthier believes that sound reasoning—which, for him (unlike Kantians), is restricted to instrumental reasoning designed to maximize self-interest—will also yield a set of common moral duties. This is because each of us is likeliest to do the best by adhering to (and acquiring a disposition to conscientiously adhere to) mutually beneficial rules that sometimes mandate self-sacrifice, and these are just what Gauthier thinks moral rules are.

Smith’s view is a kind of ideal observer theory. What we have reason to do is a matter of what an ideal adviser would want us to do. This ideal advisor shares our fundamental desires, but is fully factually informed, has flawless

reasoning abilities, and is possessed of a fully coherent set of desires. Though such advisers obviously do not really exist, they play a crucial role in determining the content of our moral duties. Smith endorses the Kantian tenet that sound reasoning is more than self-interested instrumental reasoning. He endorses the Humean view that all of our fundamental commitments are contingent. But he believes that sound reasoning on the basis of disparate commitments will nevertheless yield a universal set of rationally authoritative moral requirements: The ideal advisers would converge in all of their basic recommendations.

What all of these thinkers share is a belief that the content of our basic moral duties conceptually depends very importantly on the outcome of sound practical reasoning. Perhaps the nature of this dependence can best be seen by contrasting their metaphysical thesis with that of moral realism. Moral realists, some of whom consider themselves rationalists because of their endorsement of the epistemological and normative theses, nevertheless deny the metaphysical thesis. Whichever picture of the ideal reasoner and the ideal reasoning process we adopt, the metaphysical thesis tells us that something is our moral duty just because a process of sound reasoning would ratify it. Moral realists, by contrast, might allow for the possibility of inerrant reasoners who infallibly identify all moral duties, but realists will insist that the content of our duties is fixed in a way that does not depend on the outcomes of such reasoning. Ideal reasoners will discover the moral truths there are to be discovered, rather than creating, through the exercise of sound practical reasoning, the content of our basic moral duties.

Though moral realists and those who endorse practical reason approaches to ethics are divided in their opinion of the metaphysical thesis, those among them who accept the rationalist moniker may still agree on the epistemological thesis. Moral rationalists in this domain believe that at least some moral knowledge is a priori. While experience may be needed to obtain an understanding of moral concepts, and empirical premises might be needed to determine what our all-things-considered duty in a given case might be, knowledge of fundamental moral principles can be had without relying on contingent, empirical premises for evidential support. Rationalists here believe, for instance, that one needs nothing other than a sound understanding of the proposition to be justified in thinking that, *prima facie*, it is immoral to rape or torture people solely for personal enjoyment.

Both moral realists and adherents of practical reason approaches to ethics are divided on the availability of a

priori moral knowledge. Among moral realists, the sides are drawn by reference to whether the emerging view takes the form of ethical naturalism or ethical nonnaturalism. Naturalists will see all knowledge, and so all moral knowledge, as a posteriori, available only by applying the same methods of discovery and justification as are utilized in the natural sciences. Ethical nonnaturalists, by contrast, will reject the attempted assimilation of ethics to natural science, and so will leave room for the possibility of a priori moral knowledge. Indeed, the historically most prominent moral realists, up to and including the British intuitionists of the early twentieth century (Henry Sidgwick, G. E. Moore, and W. D. Ross), all endorsed the idea that fundamental moral principles were self-evident: knowable solely on the basis of adequately understanding the content of the relevant principle. Robert Audi, the contemporary philosopher who has done most to develop this view, agrees with his historical forbears in attributing to reason the power to discern certain fundamental truths about reality—in this case, moral reality.

Like moral realists, adherents of practical reason approaches to ethics (and so of the rationalists' metaphysical thesis) disagree among themselves about the tenability of a priori moral knowledge. Kantians, who have championed the possibility of synthetic a priori knowledge in other areas, likewise endorse its possibility in ethics. But others, such as Gauthier, are avowed naturalists who believe that knowledge of an agent's moral duties depends crucially on knowledge of her contingent, fundamental commitments, and so cannot be a priori.

The normative thesis, which states that moral requirements entail excellent reasons for action, is accepted by all practical reason approaches to ethics. Indeed, a major attraction of such theories is that they are able to explain the intimate connection that might obtain between one's moral duty and one's reasons for action. We will always have reason to do as morality says, because morality is constructed from our own deepest commitments. If moral duties are rational extensions of our fundamental commitments, then, to the extent that we are rational, we cannot be alienated from what morality requires of us.

Moral realists have had a more difficult time explaining the reason-giving power of morality. Indeed, some, such as Peter Railton and David Brink, have given up on the idea that there is any necessary connection between moral demands and reasons for action. Such thinkers believe that something qualifies as a practical reason only if it furthers an agent's ends (all of which, in their view, are contingent). They believe that moral demands may

fail to serve an agent's ends. It follows that there may be no good reason to abide by the demands of morality.

If one retains this popular view of reasons, then the moral rationalist's normative thesis can be sustained in only one of two ways. The first involves rejecting moral realism, and taking up a practical reason approach that identifies our moral duties with the rational extensions of our commitments. Here, both our reasons and our moral duties will rely on these commitments, and this can explain the perfect alignment of morality and reasons for action. The second, more controversial way of defending both rational egoism and the normativity thesis, is to accept moral realism, and then to insist that fulfillment of our moral duties will always, of necessity, further our ends. Plato argued this way, and some theists who are moral realists do so as well. This is the less traveled path, however, because its defense requires that we posit a set of objective human ends that are invariably furthered by moral conduct. Such ends might be renounced by apparently coherent and rational individuals, and this has led to a great deal of suspicion about their existence.

The more common strategy for moral realists who want to vindicate the normativity thesis is to reject the view that reasons for action must always further the agent's ends. A misanthrope, for instance, might be morally bound to rescue another, if he can do so at little or no inconvenience to himself. According to the normativity thesis, there is therefore excellent reason for him to do so, even if none of the misanthrope's ends are furthered as a result. Given who the person is, he will reject the existence of such a reason. It will seem to the misanthrope an alien demand, of only spurious rational authority. It will play no role in explaining the actions he undertakes. Still, says the realist who embraces the normativity thesis, the misanthrope's reactions in such a case do not immunize him either from the moral duty, or from the practical reason that he ignores in his neglect of the victim he might have aided.

An adherent of the practical reason approach to ethics will insist that reasons be able to engage agents who are reasoning well from their fundamental commitments. And realists cannot secure this guarantee, since they do not make the content of moral demands dependent on the outcomes of sound practical reasoning. Realists will either concede the point, as Railton and Brink do, or affirm their allegiance to the normative thesis. In the latter case, they will deny that morality must engage all who are able to reason efficiently about securing their ends. The plausibility of such a denial is the subject of much contemporary metaethical discussion.



*See also* Categorical Imperatives; Error Theory of Ethics; Ethical Subjectivism; Ideal Observer Theories of Ethics; Internalism and Externalism in Ethics; Intuitionism, Ethical; Kantian Ethics; Moral Realism; Practical Reason; Response-Dependence Theories.

### **Bibliography**

- Audi, Robert. *The Good in the Right: A Theory of Intuition and Intrinsic Value*. Princeton, NJ: Princeton University Press, 2004.
- Brink, David. *Moral Realism and the Foundations of Ethics*. New York: Cambridge University Press, 1989.
- Gauthier, David. *Morals by Agreement*. New York: Oxford University Press, 1986.
- Korsgaard, Christine. *The Sources of Normativity*. New York: Cambridge University Press, 1996.
- Railton, Peter. *Facts, Values, and Norms: Essays toward a Morality of Consequence*. New York: Cambridge University Press, 2003.
- Smith, Michael. *The Moral Problem*. New York: Oxford University Press, 1994.

*Russ Shafer-Landau (2005)*

## RATIONALITY

Philosophers have, at least characteristically, aspired to possess “rationality” but have not thereby sought exactly the same thing. Portrayed vaguely, rationality is reasonableness, but not all philosophers take rationality as dependent on reasons; nor do all philosophers have a common understanding of reasons or of reasonableness. Some theorists consider rationality to obtain in cases that lack countervailing reasons against what has rationality; they thus countenance rationality as, in effect, a default status. In ordinary parlance, persons can have rationality; so, too, can beliefs, desires, intentions, and actions, among other things. The rationality appropriate to action is practical, whereas that characteristic of beliefs is, in the language of some philosophers, theoretical.

Many philosophers deem rationality as instrumental, as goal oriented. You have rationality, according to some of these philosophers, in virtue of doing your best, or at least doing what you appropriately think adequate, to achieve your goals. If ultimate goals are not themselves subject to assessments of rationality, then rationality is purely instrumental, in a manner associated with David Hume’s position. Rationality, according to this view, is a minister without portfolio; it does not require any particular substantive goals of its own but consists rather in the proper pursuit of one’s ultimate goals, whatever those goals happen to be. Many decision-theoretic and eco-

nomical approaches to rationality are purely instrumentalist. If, however, ultimate goals are susceptible to rational assessment, as an Aristotelian tradition and a Kantian tradition maintain, then rationality is not purely instrumental. The latter two traditions regard certain rather specific (kinds of) goals, such as human well-being, as essential to rationality. Their substantialist approach to rationality lost considerable influence, however, with the rise of modern decision theory.

When relevant goals concern the acquisition of truth and the avoidance of falsehood, so-called epistemic rationality is at issue. Otherwise, some species of nonepistemic rationality is under consideration. One might individuate species of nonepistemic rationality by the kind of goal at hand; moral, prudential, political, economic, aesthetic, or some other. Some philosophers have invoked rationality “all things considered” to resolve conflicts arising from competing desires or species of rationality; even so, there are various approaches to rationality “all things considered” in circulation. The standards of rationality are not uniformly epistemic, then, but epistemic rationality can play a role even in what some call nonepistemic rationality. Regarding economic rationality, for instance, a person seeking such rationality will, at least under ordinary conditions, aspire to epistemically rational beliefs concerning what will achieve the relevant economic goals. Similar points apply to other species of nonepistemic rationality. A comprehensive account of rationality will characterize epistemic and nonepistemic rationality, as well as corresponding kinds of irrationality (e.g., weakness of will).

Taking rationality as deontological, some philosophers characterize rationality in terms of what is rationally obligatory and what is merely rationally permissible. If an action, for instance, is rationally obligatory, then one’s failing to perform it will be irrational. Other philosophers opt for a nondeontological evaluative conception of rationality that concerns what is good (but not necessarily obligatory) from a certain evaluative standpoint. Some of the latter philosophers worry that, if beliefs and intentions are not voluntary, then they cannot be obligatory. Still other philosophers understand rationality in terms of what is praiseworthy, rather than blameworthy, from a certain evaluative standpoint. The familiar distinction between obligation, goodness, and praiseworthiness thus underlies three very general approaches to rationality.

Following Henry Sidgwick, William Frankena has distinguished four conceptions of rationality: (1) an egoistic conception implying that it is rational for one to be

or do something if and only if this is conducive to one's own greatest happiness (e.g., one's own greatest pleasure or desire satisfaction); (2) a perfectionist conception entailing that it is rational for one to be or do something if and only if this is a means to or a part of one's moral or nonmoral perfection; (3) a utilitarian conception implying that it is rational for one to be or do something if and only if this is conducive to the greatest general good or welfare; and (4) an intuitionist conception implying that it is rational for one to be or do something if and only if this conforms to self-evident truths, intuited by reason, concerning what is appropriate. The history of philosophy represents, not only these conceptions of rationality, but also modified conceptions adding further necessary or sufficient conditions to one of (1)–(4).

Given an egoistic conception of rationality, one's being rational will allow for one's being immoral, if morality requires that one not give primacy to oneself over other people. Rationality and morality can then conflict. Such conflict is less obvious on a utilitarian conception of rationality. In fact, if morality is itself utilitarian in the way specified (as many philosophers hold), a utilitarian conception of rationality will disallow rational immorality. A perfectionist conception of rationality will preclude rational immorality only if the relevant perfection must be moral rather than nonmoral; achieving nonmoral perfection will, of course, not guarantee morality. As for an intuitionist conception of rationality, if the relevant self-evident truths do not concern what is morally appropriate, then rational immorality will be possible. An intuitionist conception will bar conflict between rationality and morality only if it requires conformity to all the self-evident truths about what is morally appropriate that are relevant to a situation or person. So, whether rationality and morality can conflict will depend, naturally enough, on the exact requirements of the conception of rationality at issue.

Richard Brandt has suggested that talk of what it would be rational to do functions to guide action by both recommending action and by making a normative claim that evaluates the available action relative to a standard. An important issue concerns what kind of strategy of using information to choose actions will enable one to achieve relevant goals as effectively as any other available strategy. Brandt has offered a distinctive constraint on such a strategy: A rational decision maker's preferences must be able to survive their being subjected to repeated vivid reflection on all relevant facts, including facts of logic. This constraint suggests what may be called (5) a relevant-information conception of rationality: Rational-

ity is a matter of what would survive scrutiny by all relevant information.

A relevant-information conception of rationality depends, first, on a clear account of precisely when information is relevant and, second, on an account of why obviously irrational desires cannot survive scrutiny by all relevant information. Evidently, one could have a desire caused by obviously false beliefs arising just from wishful thinking, and this desire could survive a process of scrutiny by all relevant information where the underlying false beliefs are corrected. In any case, a relevant-information conception of rationality will preclude rational immorality only if it demands conformity to all relevant moral information.

The egoistic, perfectionist, utilitarian, and relevant-information conceptions of rationality are nonevidential in that they do not require one's having evidence that something is conducive to self-satisfaction, perfection, general welfare, or support from all relevant information. Many philosophers would thus fault those conceptions as insufficiently sensitive to the role of relevant evidence in rationality. If relevant evidence concerns epistemic rationality, we again see the apparent bearing of epistemic rationality on rationality in general. The latter bearing deserves more attention in contemporary work on nonepistemic rationality.

Philosophers currently divide over internalism and externalism about rationality. If rationality demands reasons of some sort or other, the dispute concerns two senses of talk of a person's having a reason to perform an action. An internalist construal of this talk implies that the person has some motive that will be advanced by the action. An externalist construal, in contrast, does not require that the person have a motive to be advanced by the action. Bernard Williams, among others, has suggested that any genuine reason for one's action must contribute to an explanation of one's action and that such a contribution to explanation must be a motivation for the action. He concludes that externalism about rationality is false, on the ground that external reasons do not contribute to explanation of action in the required manner. Externalism about rationality does allow that reasons fail to motivate, but this, according to externalists, is no defect whatever. Externalists distinguish between merely motivating reasons and justifying reasons, contending that only the latter are appropriate to rationality understood normatively; what is merely motivating in one's psychological set, in any case, need not be justifying. Perhaps, then, disputes between internalists and externalists

will benefit from attention to the distinction between justifying and merely motivating reasons.

Modern decision theory assumes that, in satisfying certain consistency and completeness requirements, a person's preferences toward the possible outcomes of available actions will determine, at least in part, what actions are rational for that person by determining the personal utility of outcomes of those actions. In rational decision making under certainty one definitely knows the outcomes of available actions. In decision making under risk one can assign only various definite probabilities less than 1 to the outcomes of available actions. (Bayesians assume that the relevant probabilities are subjective in that they are determined by a decision maker's beliefs.) In decision making under uncertainty one lacks information about relevant states of the world and hence cannot assign even definite probabilities to the outcomes of available actions. Acknowledging that rationality is purely instrumental (and thus that even Adolf Hitler's Nazi objectives are not necessarily rationally flawed), Herbert Simon has faulted modern decision theory on the ground that humans rarely have available the facts, consistent preferences, and reasoning power required by standard decision theory. He contends that human rationality is "bounded" in that it does not require utility maximization or even consistency. Rather, it requires the application of a certain range of personal values (or preferences) to resolve fairly specific problems one faces, in a way that is satisfactory, rather than optimal, for one. Simon thus relies on actual human limitations to constrain his account of rationality.

Contemporary theorists divide over the significance of human psychological limitations for an account of rationality. The controversy turns on how idealized principles for rationality should be. This raises the important issue of what exactly makes some principles of rationality true and others false. If principles of rationality are not just stipulative definitions, this issue merits more attention from philosophers than it has received. Neglect of this metaphilosophical issue leaves the theory of rationality as a subject of ongoing philosophical controversy.

**See also** Aristotle; Bayes, Bayes' Theorem, Bayesian Approach to Philosophy of Science; Decision Theory; Hume, David; Kant, Immanuel; Sidgwick, Henry; Utilitarianism.

### **Bibliography**

Audi, Robert. *Action, Intention and Reason*. Ithaca, NY: Cornell University Press, 1993.

- Audi, Robert. *The Architecture of Reason: The Structure and Substance of Rationality*. New York: Oxford University Press, 2001.
- Benn, S. I., and G. W. Mortimore, eds. *Rationality and the Social Sciences*. London: Routledge and Paul, 1976.
- Brandt, R. A. *A Theory of the Good and the Right*. Oxford: Clarendon Press, 1979.
- Brown, Harold. *Rationality*. London: Routledge, 1988.
- Elster, J., ed. *Rational Choice*. New York: New York University Press, 1986.
- Foley, Richard. *The Theory of Epistemic Rationality*. Cambridge, MA: Harvard University Press, 1987.
- Frankena, W. "Concepts of Rational Action in the History of Ethics." *Social Theory and Practice* 9 (1983): 165–197.
- Fumerton, Richard. "Audi on Rationality: Background Beliefs, Arational Enjoyment, and the Rationality of Altruism." *Philosophy and Phenomenological Research* 67 (2003): 188–193.
- Gauthier, D. *Morals by Agreement*. Oxford: Clarendon Press, 1986.
- Harman, Gilbert. *Change in View: Principle of Reasoning*. Cambridge, MA: MIT Press, 1986.
- Hollis, M., and S. Lukes, eds. *Rationality and Relativism*. Cambridge, MA: MIT Press, 1982.
- Jeffrey, Richard. *The Logic of Decision*. 2nd ed. Chicago: University of Chicago Press, 1983.
- Katz, Jerrold. *Realistic Rationalism*. Boston: MIT Press, 1998.
- Mele, A. *Irrationality*. New York: Oxford University Press, 1987.
- Mele, A., and P. Rawlings, ed. *The Oxford Handbook of Rationality*. Oxford: Oxford University Press, 2004.
- Moser, P., ed. *Rationality in Action*. Cambridge, U.K.: Cambridge University Press, 1990.
- Nozick, R. *The Nature of Rationality*. Princeton, NJ: Princeton University Press, 1993.
- Rescher, N. *Rationality*. Oxford: Clarendon Press, 1988.
- Sidgwick, H. *The Methods of Ethics*. 7th ed. London, 1907.
- Simon, Herbert. *Reason in Human Affairs*. Stanford, CA: Stanford University Press, 1981.
- Skyrms, Brian. *The Dynamics of Rational Deliberation*. Cambridge, MA: Harvard University Press, 1990.
- Slote, M. *Beyond Optimizing: A Study of Rational Choice*. Cambridge, MA: Harvard University Press, 1989.
- Williams, B. "Internal and External Reasons." In *Rational Action*, edited by R. Harrison. Cambridge, U.K.: Cambridge University Press, 1979.

**Paul K. Moser (1996)**

*Bibliography updated by Benjamin Fiedor (2005)*

## RAVAISSON-MOLLIEN, JEAN GASPARD FÉLIX (1813–1900)

The French spiritualist philosopher and art historian Jean Gaspard Félix Ravaisson-Mollien was born in Namur, Belgium. He received his philosophical training in Munich under Friedrich von Schelling and took a degree

in Paris in 1838 under Victor Cousin. His philosophical work began with his prize essay, *Essai sur la métaphysique d'Aristote*, and a short teaching career at Rennes in 1838. In 1840 he was appointed inspector general of libraries, a post that he held until 1860, when he became inspector general in the department of higher education. Meanwhile, as a semiprofessional painter he had become interested in classical antiquities, and in 1870 he was made curator in the department of antiquities in the Louvre. The fruit of this was his well-known set of reconstructions of the Venus de Milo.

The most influential of Ravaisson's publications was his *Rapport sur la philosophie en France au XIX<sup>e</sup> siècle*, made at the request of the imperial government in 1867. At this time the school of Cousin was in the ascendancy in France, and it was difficult, indeed practically impossible, for a man who was not an eclectic to get an appointment in the university system. Ravaisson's purpose in his report was to show that there was a continuity in the French philosophical tradition and that French philosophers had always presupposed metaphysical principles that implied what he called spiritualism. This tradition, he maintained, always swung between sensationalism, phenomenalism, and materialism, on the one hand, and idealism, on the other. But spiritualism really began in the nineteenth century with Maine de Biran, who used as his starting point the human will and who held that the will is independent both of sensations and of ideas. This viewpoint, Ravaisson argued, was not only the proper beginning of a philosophy but also the only one that could unify the opposing tendencies of empiricism and idealism.

Such a conclusion was in clear contradiction to the tenets of Cousin's eclecticism, which aspired to fuse "the best in each philosopher." Ravaisson tried to show that such a fusion in reality consists in refuting those philosophies which displease the eclectic and retaining those which please him. In classifying all philosophies under the headings of sensualism, idealism, mysticism, and skepticism, Cousin accepted only that philosophy which he called idealism but which, said Ravaisson, was really a simple mixture of the Scottish philosophy of common sense with a few ideas from Maine de Biran. The eclectics, moreover, failed to understand these ideas. Ravaisson claimed for himself the credit of introducing the true thought of Maine de Biran to his contemporaries. Readers of this report were thus informed that the de facto official philosophy of the French universities was not only a foreign importation but also untrue.

Ravaisson was not satisfied with undermining eclecticism. He also felt it important to point out the weaknesses of positivism. These weaknesses, he claimed, arose from the identification of philosophical method with the methods of science. Science, which admittedly studies the external world, can never tell us anything about the internal world of thoughts, aspirations, desires, and dreams; and when it attempts to do so, it transforms them into quasi-external objects. This inevitably leads to materialism, for the laws of matter are the only laws that science can formulate. Science's basic categories are space and quantity, and its basic method is analysis. But the phenomena of consciousness are never spatial or quantitative, and to attempt to categorize them in these terms is to change their essential nature.

Ravaisson's report reviewed all the contemporary schools of thought and all the contemporary philosophers. It was a model of patience and thorough investigation and has become the primary source of information about individuals who are obscure and in some cases forgotten. It did not stop at professional philosophers but looked into the presuppositions of scientists, such as the physiologist Claude Bernard and the psychiatrist Albert Lemoine. In every case, Ravaisson found either too strong an emphasis on the dependence of the "spirit" upon material causes or an identification of ideas with strictly logical, hence analytical, reason. Whereas one set of philosophers tried to explain the mind in terms that were inappropriate, the other failed to ask the central question of why the mind operated as it did. Neither group could explain our undeniable feeling of being active causes; neither could see why the spirit needs both analysis and synthesis.

Whether the object of our thinking is the external or the internal world, it will be found that we have to use two absolutely general metaphysical principles, that of an infinite reality and that of limitation. The dialectical reason for this is that every analytical sentence distinguishes between parts of a whole and no whole can be discussed except by reference to its constituent parts. But Ravaisson did not rest his doctrine on this dialectical argument. On the contrary, he believed that history had shown that every philosopher presupposes these principles, whether he knows it or not. The tendency of the history of philosophy is toward the progressive realization of this truth. It is implicit in all philosophy and is steadily becoming explicit. Ravaisson's report thus presented not only an exposition of contemporary French philosophies but also a theory about the history of philosophy.

In a shorter study, *De l'habitude*, written as a thesis at the Sorbonne in 1838, Ravaissou returned to the problem raised in Maine de Biran's prize-winning essay on the influence of habit on thinking. Ravaissou's study is of special historical interest since it forms the nucleus of the philosophy of Henri Bergson.

At the beginning of his argument Ravaissou laid down a fundamental distinction between the roles played by space and time in our lives. "Space," he said, "is the most obvious and elementary condition and form of stability or permanence; time the universal condition of change." Corresponding to these two basic principles are matter and life respectively. In matter there is no individuality and no possibility of habit, a point that Ravaissou probably encountered in his study of Aristotle. Life, on the contrary, forms a world of its own, a world that is internal to the living being. A set of oppositions follows, that of necessity (matter) versus that of "nature" (life), a set that echoes the two realms of necessity and freedom elaborated by Schelling. The repetition of a change modifies "nature," and the living being swings between the limitations of its material conditions and its own inner freedom. As the forms of life develop, their power of spontaneous action becomes greater, so that although the inorganic is timeless, life implies a "definite continuous *durée*." As we move up from vegetable to animal to human life, we find that whereas sensory impressions become weaker when repeated, our powers of movement become stronger and stronger.

Corresponding to these dualities is another. Within the human soul are the two powers of understanding and of activity. The understanding sees everything under the aspects of diversity, quantity, and space; the power of activity appears primarily in our feeling of effort, which is gradually reduced by habit. Habit transforms voluntary movements into instinctive movements. Voluntary movements could not be made if there were no resistance from without, but for them to be made at all requires that somewhere there be an undetermined center of activity, which is the will. And when one asks what the will is seeking, the answer is that it seeks the good, or God. It is not difficult to see in these views both the influence of Schelling and the anticipation of Bergson.

**See also** Aristotle; Bergson, Henri; Bernard, Claude; Cousin, Victor; Empiricism; Idealism; Maine de Biran; Materialism; Mysticism, Nature and Assessment of; Phenomenalism; Positivism; Schelling, Friedrich Wilhelm Joseph von; Sensationalism; Skepticism; Space; Time.

## Bibliography

### WORKS BY RAVAISSOU

*Essai sur la métaphysique d'Aristote*, 2 vols. Paris, 1837–1846.

*De l'habitude*. Paris: H. Fournier, 1838. Reprinted in *Revue de métaphysique et de morale* 2 (1894): 1–35.

*Rapport sur la philosophie en France au XIX<sup>e</sup> siècle*. Paris, 1867.

*Testament philosophique et fragments*. Paris, n.d. Published posthumously, with bibliography.

### WORKS ON RAVAISSOU

Bergson, Henri. "Notice sur la vie et les oeuvres de M. Félix Ravaissou-Mollien." *Comptes-rendus de l'Académie des sciences morales et politiques* (1904). Reprinted in Bergson's *La pensée et le mouvant* (Paris: Alcan, 1934) as well as in the *Testament philosophique* of Ravaissou.

Dopp, Joseph. *Félix Ravaissou, la formation de sa pensée d'après des documents inédits*. Louvain: Editions de l'Institut Supérieur de Philosophie, 1933.

Gunn, J. Alexander. *Modern French Philosophy*, 73–75 and passim. London: Unwin, 1922.

George Boas (1967)

## RAWLS, JOHN

(1921–2002)

John Rawls is widely regarded as one of the most significant political philosophers of the twentieth century. Educated at Princeton University, he taught at Cornell University and the Massachusetts Institute of Technology before joining the faculty of Harvard University in 1962. Rawls's *A Theory of Justice* (1971) revitalized political theory as an academic discipline and rejuvenated interest in the substantive social issues that had long been neglected by academic philosophers. Rawls continued to refine and defend his theory in a series of articles and lectures, the most important of which he revised and collected in his 1993 work *Political Liberalism*. In 1999 *The Law of Peoples* extended his theory to questions of international relations, and in the next two years, despite declining health, he published *Lectures of the History of Moral Philosophy* (2000) and *Justice as Fairness: A Restatement* (2001).

### JUSTICE AS FAIRNESS

The primary objective of Rawls's political theory is to articulate and defend a conception of justice for a modern democratic regime. The theory begins with the idea of society as a fair system of cooperation between free and equal persons. The principles of justice for such a society characterize its fair terms of cooperation by specifying its citizens' basic rights and duties and by regulating the distribution of its economic benefits. To formulate his particular conception of justice, Rawls

invokes the familiar theory of the social contract, according to which the legitimate rules for a society are arrived at by the autonomous agreement of its members. Rawls's version of the contract theory is distinctive, however, in its insistence on the essential fairness of the point of view from which the agreement itself is conceived. This enables Rawls to appeal to the justificatory force of *pure procedural justice*, the idea that the fundamental fairness of a procedure can ensure the justice of its outcome provided that there is no independent criterion for the justice of that result. Fairness thus characterizes both the terms of the contractual agreement and the conditions in which that agreement is made. Rawls appropriately names the resulting theory *justice as fairness*.

Rawls's contractarian or *constructivist* theory represents this fundamental ideal of fairness by situating the contracting parties in a hypothetical *original position*. The most important feature of this theoretical model is the *veil of ignorance*, which denies to the parties any knowledge of their actual natural endowments, their social position, or even their conception of what makes for a good life. As a consequence, the parties cannot determine how proposed principles would affect their interests personally. The veil of ignorance thereby reflects our conviction that it would be patently unreasonable to allow principles that favored any individuals or groups merely in virtue of their possession of morally arbitrary attributes such as their race or sex, or because they happened to affirm a particular religious or philosophical doctrine.

## THE PRINCIPLES OF JUSTICE

Though deprived of knowledge of their particular ends, attachments, and aspirations, the parties in the original position are still rationally motivated to further their conception of the good, whatever it is. They also have *higher-order interests* in developing and exercising the two *moral powers* that they share as free and equal beings: (1) the capacity to understand and act from a sense of justice and (2) the capacity to form, revise, and rationally pursue a conception of the good. The parties will therefore seek for themselves the best possible package of *primary goods*, those all-purpose, socially regulable opportunities and resources needed to advance those interests. Rawls's enumeration of these primary goods includes basic rights and liberties, the powers and prerogatives of offices and positions, income and wealth, and the social bases of self-respect. Assuming that a society has reached a minimal level of economic development, Rawls argues that the following two principles for allocating the primary goods would be selected:

- a. Each person has an equal right to a fully adequate scheme of equal basic liberties which is compatible with a similar scheme of liberties for all.
- b. Social and economic inequalities are to satisfy two conditions. First, they must be attached to offices and positions open to all under conditions of fair equality of opportunity; and second, they must be to the greatest benefit of the least advantaged members of society. (Rawls 1993, p. 291)

Since the first principle is given absolute priority over the second, Rawls argues that the basic liberties guaranteed by it, such as freedom of religion or the right to run for political office, cannot be sacrificed for any amount of personal or collective economic benefit. Such liberties can be limited only to protect the central range of application of other conflicting liberties, as when the right to a fair trial necessitates some restrictions on the freedom of the press. Specific rights are included in the protection of the first principle if agents in the original position would rationally require them. For example, freedom of religion would be insisted on by the parties, for they could not risk the possibility that their religion, should they have one, would be a minority faith subject to repression by a dogmatic majority.

The second principle deals with economic and social primary goods such as income and wealth. Its second condition, the so-called difference principle, stipulates that any departures from equality of resources can be justified only if the resulting inequality benefits the least advantaged members of society. Thus, positions that require the development of talents and the expenditure of extraordinary effort might deserve greater economic rewards, but only if the increased productivity generated by such a differential would improve the condition of the least well off. Rawls argues that this requirement would be the reasonable and rational choice of individuals who, because of the fairness conditions imposed in the original position, did not know their natural and social endowments and therefore could not determine their actual position in the social order.

The first part (of the second principle) stipulates that even the limited inequalities that would satisfy the difference principle are permissible only if the positions that give rise to them are open to all under conditions of *fair equality of opportunity*. This strong requirement goes beyond mere prohibition of discrimination based on arbitrary features such as gender or race. It demands that all individuals of like natural ability and similar motivation should have the same opportunities throughout their entire lives, a requirement that obviously necessi-

tates equal access to education, health care, and other social resources.

### STABILITY

A viable political theory, Rawls insists, must be practical. The well-ordered society that it mandates must be feasible and stable given realistic economic, cultural, and psychological assumptions. In *A Theory of Justice* Rawls argues that a society regulated by justice as fairness would be stable since the laws of moral psychology show that its members would tend to acquire and maintain a common comprehensive moral doctrine that would sustain it. In *Political Liberalism*, however, he admits that a liberal, nonauthoritarian regime would be characterized by a plurality of reasonable, though incompatible comprehensive religious and moral doctrines. Nonetheless, he believes that the requirement of stability can be met by justice as fairness if it is understood as a political theory. As such, it regulates only the *basic structure* of society: the background institutions that specify political and civil rights and that determine entitlements to other socially regulated goods. Members of a well-ordered society may therefore hold deeply conflicting comprehensive religious and moral views, yet still endorse a common political conception of justice as the focus of an *overlapping consensus*. Moreover, Rawls stresses that this consensus can be more than a mere *modus vivendi*, a practical compromise based on a tenuous balance of power. Rather, it can express a genuine moral commitment that reflects ideas and values implicit in the society's political culture, such as its conception of the citizen as a free and equal person and its willingness to rely on reasonable standards of *public reason* in the conduct of its political affairs.

### THE LAW OF PEOPLES

Rawls applies his principles of justice to individual, self-contained societies and not to humanity at large. He does not think, for example, that the difference principle should be applied globally. Rather, he argues for a *law of peoples*, a more limited set of obligations on just societies, first, to obey some traditional canons of international law (such as to wage war only in self-defense and to honor basic human rights) and, second, to aid peoples that lack sufficient resources to support just social institutions. These duties, he argues, would be agreed to in a second original position, populated now by representatives of just or "decent" peoples who are behind a veil of ignorance with respect to the particular societies that they represent.

### REFLECTIVE EQUILIBRIUM

Rawls's methodology has been as influential and as controversial as his substantive views. Declining to ground his views on any deep metaphysical or other philosophical truths, Rawls maintains that political theory should formulate a coherent set of principles that accounts for the considered convictions that we actually hold. The process goes beyond mere summarization of particular considered judgments, however, for it also postulates theoretical models, mediating ideas, and principles at all levels of generality. All judgments and principles are held open to revision in light of other aspects of the theory, until no further changes are needed to develop a compelling and coherent view. The resulting theory is then said to be in *reflective equilibrium*. It is also *objective*, Rawls contends, because it would gain the assent of all reasonable individuals on due reflection.

**See also** Justice; Good, The; Liberty; Moral Psychology; Political Philosophy, History of; Political Philosophy, Nature of; Rights; Social and Political Philosophy; Social Contract.

### Bibliography

#### WORKS BY RAWLS

*A Theory of Justice*. Cambridge, MA: Belknap Press of Harvard University Press, 1971. A somewhat revised 1975 version has been translated into several languages.

*Political Liberalism*. New York: Columbia University Press, 1993. Contains revised versions of several important papers published between 1974 and 1993. The revised paperback edition of 1996 has an added preface.

*The Law of Peoples*. Cambridge, MA: Harvard University Press, 1999.

*Collected Papers*, edited by Samuel Freeman. Cambridge, MA: Harvard University Press, 1999.

*Lectures on the History of Moral Philosophy*. Edited by Barbara Herman. Cambridge, MA: Harvard University Press, 2000.

*Justice as Fairness: A Restatement*, edited by Erin Kelly. Cambridge, MA: Harvard University Press, 2001.

#### WORKS ABOUT RAWLS

Arneson, Richard J., et al. "Symposium on Rawlsian Theory of Justice: Recent Developments." *Ethics* 99 (4) (1989): 695–710.

Barry, Brian. "The Liberal Theory of Justice: A Critical Examination of the Principal Doctrines." In *A Theory of Justice*, by John Rawls. Oxford, U.K.: Clarendon Press, 1973.

Blocker, H. Gene, and Elizabeth Smith, eds. *John Rawls' Theory of Social Justice: An Introduction*. Athens: Ohio University Press, 1980.

Daniels, Norman, ed. *Reading Rawls: Critical Studies on Rawls' Theory of Justice*. New York: Basic Book, 1975.

Freeman, Samuel, ed. *The Cambridge Companion to Rawls*. New York: Cambridge University Press, 2003. An

outstanding series of original articles and an exhaustive bibliography.

- Kukathas, Chandran, and Philip Pettit. *Rawls: A Theory of Justice and Its Critics*. Cambridge, U.K.: Polity University Press, 1990.
- Martin, Rex. *Rawls and Rights*. Lawrence: University Press of Kansas, 1985.
- Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books, 1974.
- Pogge, Thomas W. *Realizing Rawls*. Ithaca, NY: Cornell University Press, 1989.
- Richardson, Henry, and Paul Weithman, eds. *The Philosophy of Rawls: A Collection of Essays*. 5 vols. New York: Garland, 1999.
- Sandel, Michael J. *Liberalism and the Limits of Justice*. New York: Cambridge University Press, 1982.
- Wellbank, J. H., Denis Snook, and David T. Mason. *John Rawls and His Critics: An Annotated Bibliography*. New York: Garland, 1982.

*Alan Fuchs (1996, 2005)*

## REALE, MIGUEL

(1910–)

Miguel Reale, the Brazilian philosopher of law, historian of ideas, and politician, was a professor of law and a rector at the University of São Paulo, where he founded the Instituto Brasileiro de Filosofia and its journal, *Revista brasileira de filosofia*. Reale is a prolific author, and his books embrace the full range of his concerns, although his greatest contribution lies in the philosophy of law.

Reale has developed an analytical method (derived from German phenomenology and Italian historicism) that he calls “critical ontognoseological historicism.” Rejecting both traditional realism and idealism, he locates the transcendental conditions of human experience and knowledge in a fundamental and inseparable correlation of subject and object. These conditions are mutually implicit and reciprocally necessary and are comprehensible only as moments in a polar dialectical process. Man’s being emerges only through his own historicity, as values are realized in time through his conduct. The person finds his essence (*ser*) in what he ought to be (*dever-ser*), and he is the source of all values. Values are possible only where there are persons, and personality consists in conduct that is comprehensible only with reference to ends and values. A phenomenological description of human action reveals its essential orientation toward ends that represent values determining action and serving as the foundation of the “ought-to-be” in which man finds his essence. Reale interprets human history as a process through which values are converted into ends,

accompanied by cultural crises whenever a new generation refuses to recognize the value of traditional ends.

Legal phenomena are basic to the realization of values in common. In law, two persons are joined in a polar nexus of common needs. Reale distinguishes three traditional approaches to the understanding of the nature of law. Sociologism interprets law as a positive fact and explains it in sociological and historical terms. Neopositivism interprets law as the expression of the operative norms of a given society and analyzes its function therein. Culturalism interprets law as axiological in nature and investigates the transcendental conditions that make it possible. Reale rejects all three as merely partial interpretations. Fact, norm, and value, in his view, are dialectically unified and not merely juxtaposed. Law is a fact through which values are made concrete in history and through which intersubjective relations are normatively ordered.

**See also** Historicism; Idealism; Ideas; Latin American Philosophy; Philosophy of Law, History of; Philosophy of Law, Problems of; Realism; Value and Valuation.

### **Bibliography**

#### WORKS BY REALE

- Teoria do direito e do estado* (Theory of law and the state). São Paulo: Martins, 1972.
- Filosofia do direito* (Philosophy of law). São Paulo: Edição Saraiva, 1962.
- Pluralismo e liberdade* (Pluralism and liberty). São Paulo: Edição Saraiva, 1963.

#### WORKS ON REALE

- Revista brasileira de filosofia* 11, Fasc. 42 (April–June 1961). Dedicated to Reale on the occasion of his fiftieth birthday, this issue contains five expository articles as well as other material about him, his thought, and activity.

*Fred Gillette Sturm (1967)*

## REALISM

In the early history of philosophy, particularly in medieval thought, the term *realism* was used, in opposition to nominalism, for the doctrine that universals have a real, objective existence. In modern philosophy, however, it is used for the view that material objects exist externally to us and independently of our sense experience. Realism is thus opposed to idealism, which holds that no such material objects or external realities exist apart from our knowledge or consciousness of them, the whole universe thus being dependent on the mind or in some sense mental. It also clashes with phenomenalism,



which, while avoiding much idealist metaphysics, would deny that material objects exist except as groups or sequences of *sensa*, actual and possible.

### THE POLEMIC AGAINST IDEALISM

At the close of the nineteenth century, idealism was the dominant Western philosophy, but with the opening of the twentieth century, there was an upsurge of realism in Britain and North America, associated in the former with G. E. Moore, Bertrand Russell, and Samuel Alexander and in the latter with William James (despite his pragmatism), the new realists, and later the critical realists. Before a discussion of realist doctrine, a brief survey may be given of its attack on idealism.

The claim that material objects cannot exist independently of mind had been made on various grounds. First, the analysis of perception, especially of illusions, was held to show that our knowledge was limited to groups of sensations “in the mind” or to products of the synthesis or interpretation of sensory data. Later idealists, under the slogan “all cognition is judgment,” stressed the role of judgment and interpretation in perception, concluding that objects as we know them must be largely or even wholly the work of the mind. Second, physical objects cannot exist independently of the mind, for whatever is known is relative to the mind that knows it. This is the “egocentric predicament”—that one can never eliminate the “human mind” from knowledge and discover what things are like apart from one’s consciousness or, indeed, whether they exist when they are not known, for the discovery itself involves consciousness and thus would be knowing. This may also be stated in terms of the doctrine of internal relations—that the nature of anything is grounded in and constituted by the relations it has with other things; no two related things could be what they are if the relation between them did not exist, and so, as a special case of this, physical objects could not be as they are apart from their relation to the mind that knows them.

**STATUS OF THE OBJECTS OF PERCEPTION.** Concerning the analysis of perception, realist philosophers have devoted considerable attention to showing that in perception we obtain knowledge of external physical objects either directly or by means of *sensa*. Their accounts of perceiving and their solutions to the problems raised by illusions and other facts of perception differ greatly, but they agree in rejecting the view that things cannot exist unperceived. G. E. Moore’s influential “Refutation of Idealism” consisted in an attack on this thesis, which, follow-

ing George Berkeley, he stated as “*esse is percipi*” (“to be is to be perceived”). He claimed that in maintaining this the idealists had failed to distinguish between the act and the object in sensation. They had confused the sensation of blue with its object blue or, when claiming to distinguish them, inconsistently treated them as identical.

Sensations are alike in being acts of awareness but differ in what they are awareness of. Once the object is distinguished from the awareness of it, there is no reason to deny its existence unperceived. Further, in no other situation have we a better claim to be aware of something distinct, so that if sensations are not cases of awareness of objects, no awareness is ever awareness of anything, and we cannot be aware of other persons or even of ourselves and our own sensations. Fundamentally, Moore’s thesis concerning sensations rested on introspection; it has been denied on a similar introspective appeal by upholders of the adverbial analysis of sensing, and Moore himself later had grave doubts about it. Commonsense realists would say that he conceded too much in talking of sensations and interpreting “being perceived” (*percipi*) as “being sensed” (*sentiri*); the proper starting point is our awareness of material objects. But Moore was no doubt accepting the usual conclusions from the argument from illusion. From his analysis arises the question: “What is the object of sensation?” The answer, “A sense datum,” posed the problem, which he never solved, of the relation between sense data and material objects. It was met by others with some form of representative realism or, more usually, phenomenalism. Phenomenalism, however, particularly if coupled with the adverbial analysis of sensing, means the abandonment of realism. The idealist stress on judgment in perception was at first little discussed, but critical realism and the sense-datum theory later offered more plausible alternatives.

**THE EGOCENTRIC PREDICAMENT.** The realist attack on the egocentric predicament involved considerable discussion, particularly in the United States, and led to some close argument—for example, in attempts to show that the idealist principle led to self-contradiction or circularity when developed. The egocentric predicament was claimed to have no idealist implications. To infer from it that nothing exists outside consciousness is simply fallacious—that one cannot discover *X* does not mean that *X* does not exist or even that it is unreasonable to suppose that *X* exists. Indeed, if it were true that things could not exist apart from a person’s consciousness of them, neither, presumably, could other persons; the predicament would imply an incredible solipsism.

Nor is there any evidence of the lesser conclusion that objects outside consciousness would be quite different. No conclusion about the degree of distortion introduced by our consciousness follows from its ubiquity, and it may be negligible; one can only try to discover the degree by comparing various methods of knowing. (Distortion by the method of observation may be serious in atomic physics, but the same argument that establishes distortion there shows it to be negligible for objects larger than atoms.)

The predicament is sometimes stated in terms of the privacy of experience—a person can never know anything that is not a content of his private experience. This, however, is question-begging in that it simply denies the ordinary assumptions that we are aware of other persons and external public objects. There may be grounds for denying these assumptions in certain cases, but such grounds rest on evidence of causal processes and of illusions, evidence that is largely obtained from other persons, or with the aid of public objects, or from comparisons with perceptions of public objects. Further, though more dubiously, Wittgenstein has argued that if we had only private experiences, not only would they be incommunicable, but also we could not describe or speak about them even to ourselves, for the use of language implies rules that are communal and have to be established and checked with respect to public objects.

Against the doctrine of internal relations it was claimed that relatedness is compatible with independence, that the same thing can enter into a variety of relations without losing its identity. This seemed so obvious that James confessed to finding it “weird” to have to argue for it. (Anticipating a contemporary approach, he accused the idealists of confusing linguistic or conceptual differences with factual ones; in referring to two relations of an object, our phrases and thoughts differ, but there is no corresponding difference in the object itself.) As the realists were defending what in their eyes was obvious, they were forced into detailed criticism rather than into the kind of positive thesis that can be readily summarized.

This battle was certainly won by the realists in that few English-speaking philosophers in the twentieth century espoused idealism. Indeed, to anyone coming from contemporary discussions, the controversy has an air of unreality. Partly this is because in a climate of thought that respects common sense and science, realism seems so obvious a starting point that it is difficult to explain how the idealist view ever seemed plausible; partly it is because current idioms, issues, and logical presuppositions are so

different from earlier ones. Granted, however, that material objects exist independently of our perception, the difficulties facing a realist account of this perception still remain and cause serious divisions among realists.

## DIRECT REALISM

Direct realism is the general view that perception is a direct awareness, a straightforward confrontation (or in touch, contact) with the external object. It may be further subdivided according to the various attitudes then taken toward illusions and hallucinations. In contrast, there are the various types of indirect or dualist realism, which claim that perception is primarily of mental representations of the external object, as in traditional representative realism, or that our perception of the external object is by means of private, mental *sensa*.

**NAIVE REALISM.** Naive realism is the simplest form of direct realism and is usually alleged by philosophers to be an innocent prejudice of the average person that has to be overcome if philosophical progress is to be made. It is normally stated in terms of sensible qualities or *sensa*. When we look around us, we can distinguish various colored, shaped expanses that we suppose to be the surfaces of material objects, we may hear various sounds that we suppose to come from such objects, we may feel something smooth and hard that we suppose to be a table top, and so on. Naive realism claims that these suppositions are all correct—that the shapes, colors, sounds, and smooth, hard expanses (the sensible qualities) are always the intrinsic properties of material objects and in sight and touch are their surfaces.

Such a claim can easily be shown to be erroneous by the argument from illusion. When *A* looks at the table from above, he sees a round expanse; when *B* looks at it from a distance, he sees an elliptical one. Without self-contradiction, however, the round and elliptical shapes cannot both be the surface of the table—that is, an intrinsic property. Similarly, when *C*, who is color-blind, looks at a red book, he sees a black shape that, again, cannot be the surface of that red book; when *D*, a drunkard, sees snakelike shapes on the bed, they are not real snakes. Such examples may be multiplied indefinitely and dispose of naive realism as thus stated, but commonsense realists would say that the doctrine misrepresents the views of the average person and that philosophical discussions of it beg the question in favor of dualism by speaking of sensible qualities or *sensa* as distinct from physical objects.

NEW REALISM AND THE SELECTIVE THEORY. The new realists—E. B. Holt, W. T. Marvin, W. P. Montague, R. B. Perry, W. B. Pitkin, and E. G. Spaulding—are notable chiefly for a common realist platform published in 1910 and expanded in 1912 and for their polemic against idealism. Their realism was carried to the Platonic extreme of claiming real existence for logical and mathematical entities, and they had difficult and conflicting views about consciousness. Without, however, pursuing these, we may note their main attempt (by Holt) to deal with illusions, which is a version of what is often called the selective theory. The essential points of this theory are, first, all the various appearances of an object are its intrinsic, objective properties and are directly apprehended by the percipient. For example, the table that looks round to *A* and elliptical to *B* is intrinsically both round and elliptical; the mountain that looks green close up and blue in the distance is both green and blue. There is nothing private or mental about such appearances, for they can be photographed, as can mirror images and various optical illusions. Second, the function of the nervous system and of the causal processes in perception is to select and reveal to the percipient one property from each set of properties, for example either the elliptical or the round shape of the table.

One difficulty in this is that it does not really account for error. If we are always directly aware of actual characteristics of objects, what sense does it make to talk, as we do, of illusions, mistakes, or misperceptions? Another lies in the weakness of the selective theory compared with the generative theory, adopted by dualist realism, which states that the sensible qualities, or *sensa*, are “generated,” by the action of the object on the sense organs and nervous system and thus are not intrinsic properties of external objects. The usual reasons for preferring the generative theory are, on the one hand, that it is self-contradictory to say the table is intrinsically both round and elliptical or the mountain is intrinsically both green and blue. Furthermore, objects must be incredibly complex if they are to possess all these shapes and colors, plus, presumably, qualities corresponding to the queer appearance of objects when one has taken mescaline or suffers from giddiness or double vision. On the other hand, it is not clear how the nervous system specifically responds to or selects one of the various shapes, colors, and so on. This is particularly so in such cases as color blindness, drugs, and double vision, where the different appearances are the result of differences in the percipient and where the pattern of light waves can be detected as already differentiated for the shape and color normally perceived.

The generative theory, however, fits the facts of the causal processes quite well; it is natural to suppose that the generation of the sensory experience and its *sensum* occurs at the end of the causal chain that extends from object to brain by way of sense organ and nerves. This is confirmed by the reproduction of such experiences in mental imagery (presumably because the appropriate brain activity recurs), by the sensations resulting from electrical stimulation of the brain, and by the time lag that may occur between an event and our perception of it—all things that the selective theory cannot explain. Also, the generative theory can explain how voluntary selection occurs. When we turn our head to look at *X* rather than *Y*, we are allowing light from *X* rather than *Y* to strike our eyes and thus bring into being the *sensa* appropriate to *X*. As to photographing appearances, the photograph corresponds to the retinal image, not the *sensum*—that is, it reproduces not the perceived appearance but an intermediate cause of it; to enter into human experience, it must, in turn, be perceived by generating *sensa*.

PERSPECTIVE REALISM AND THEORIES OF APPEARING. The first objection to the selective theory—that it makes objects possess contradictory qualities—might be met by stressing that shapes, colors, and other qualities are not intrinsic but relative properties. The table is round from here, elliptical from there; the mountains are green in this light, blue in that light, and so on. This idea has been coupled with direct realism in a number of similar theories: perspective realism (E. B. McGilvary), objective relativism (A. E. Murphy), or the theory of appearing. (This last name was given by H. H. Price to a view put forward by H. A. Prichard. Roderick M. Chisholm, however, uses it more widely, and it is convenient to class all these views as theories of appearing.) Their central point is that direct realism can deal with illusions, or at least perceptual relativity, by saying that sensible qualities are not possessed by the object *simpliciter* but are always relative to some point of view or standing conditions. We always perceive sensible qualities in some perspective—spatial, even temporal (we see the distant star as it is from here and now), or illuminative (the object as it is in this light). (In such theories the shape, color, and so on are possessed by the object at its own location but are perceived subject to perspective, meaning from a viewpoint. In contrast, Bertrand Russell had a phenomenalistic theory of “perspectives” that were spread through space as possible *sensa* and actualized by or in the percipient.)

Such perspective-realist statements as “The table is round from here” sound forced, for the natural word to

use is *looks*, not *is*, and it is possible to express this kind of direct realism in terms of looking or appearing. Physical objects simply are such that they appear different from different positions, and we see them as they appear from a viewpoint or in certain conditions. Thus, we may see the round table looking elliptical from here, but even so it is still the table that we see.

Thus far the theory is trite and does little more than state the situation in a way that dualists could accept and then claim to analyze. To be distinctive, it must, as its essential characteristic, separate directness and incorrigibility. Sense-datum theory links the two, assuming that if we see an object directly, we must see it as it actually is. Thus, when the round table looks elliptical, we do not see it directly; what we see directly is an elliptical datum belonging to it.

In contrast, theories of appearing must simply claim that seeing an object directly is compatible with variation or even error in perception, so that we still see it directly when according to viewpoint, lighting, and similar factors, it appears really different from what it is. (Some might object that the theory cannot admit that perceiving is ever erroneous. Perspective realism treats all properties as relative and all perspectives as equal—the table is round from here, elliptical from there, but not round in itself; similarly all appearances should be treated as equally valid. Nevertheless, it seems more plausible to treat some appearances as privileged; in some conditions we see the real shape, the round object appearing as it is—that is, round. It may be considered a weakness of the perspective theory that it does not take into account the fact that objects do seem to have real [measured] shapes and volumes absolutely, not relative to a viewpoint.)

The approach of theories of appearing may deal plausibly with perspectival and similar variations, but it has two main defects. First, not all variations are of this nature. In double vision or mescaline illusions there seems to be existential appearing—there may appear to be two or even many tables when we look at one table. Price has argued that this cannot really be a case of directly seeing one table, for it differs significantly from seeing something merely with different properties, such as seeing a brown table instead of a black one. Also, many illusions are the result of subjective factors, so that it is difficult to say that one has a genuine perspective.

Talk of physiological perspectives is little help. “The bottle from here” is not on a par with “the bottle as it is to someone who has taken mescaline,” for mescaline may cause a range of different experiences. Similarly, when a sentry at night is convinced he sees the enemy approach-

ing but only a shadow is there, is he directly seeing the shadow in some special perspective, such as “the way it is to an anxious sentry” or “looking like a man”? Another anxious sentry might see it as a shadow and say it does not look like a man. And in a full hallucination there is no object at all. Second, theories of appearing cannot deal plausibly with the causal processes in perception since they have to adopt the selective theory. Further, we do know with varying degrees of completeness why things suffer perspectival distortion or how they cause illusion. The explanations concerned are often in terms of the causal processes and so seem to call for the generative theory and the abandonment of direct realism.

**COMMONSENSE REALISM.** In the tradition of Thomas Reid, revived by G. E. Moore, many twentieth-century British philosophers defended what they took to be a commonsense view of perception. Moore’s defense was primarily of the certainty of such simple perceptual statements as “This is a hand”; he argued that denial of these statements leads to inconsistency in beliefs and behavior and that the grounds for their denial involve propositions less certain than they are. However, his analysis of such statements in terms of sense data led away from direct realism and the commonsense view of the nature (as opposed to the reliability) of perception.

Defense of common sense became particularly associated with the Oxford linguistic analysts. Strong critics of the sense-datum theory (unlike Moore), they also reject the traditional naive realism as unfair to common sense—after all, we do not think that everything we see is the surface of a physical object (certainly not lightning flashes or rainbows) and are quite ready to admit that we often see things looking different from what they are. Although quarreling with the common philosophical uses of *appear*, *direct*, and *real*, they maintain a direct realism not unlike the theories of appearing and attempt to show in detail that in so-called illusions, including reflection and refraction, we do actually see the physical object concerned. Criticism has been made of the view that hallucinations are indistinguishable from normal perception, and more positively it may be claimed that hallucinations are mental images confused with perceptions owing to such special circumstances as drugs or fever. It is doubtful whether this can explain all the cases, and the role of the psychological processes—for example, in attention or in the influence of expectation and past experience—throws doubt on the directness of perceiving.

Some attempt has also been made to deal with the causal processes, but not very convincingly. Attacks have been made on the dualist interpretation for making it seem that we perceive something in our heads and not external objects and for the view that perceiving involves awareness of sensations. But linguistic analysts have said little of a positive nature; their main attitude is that the causal processes are at most only the conditions of perception and are the concern of the scientist but that the philosopher is concerned with perception itself, which is a skill or instantaneous achievement, not a physical process or the final stage of one. Unfortunately, scientists generally claim that the study of the causal processes requires representative realism, and even if the average person does not bother about them, an adequate philosophical theory cannot ignore the causes and conditions of perceiving, particularly since the explanation of illusions depends on them.

#### INDIRECT OR DUALIST REALISM

Many realists are persuaded by the argument from illusion and by their study of the causal and psychological processes in perception to reject direct realism and to distinguish between external material objects as the causes and ultimate objects of perceiving and private *sensa* that are the mental effects of brain processes due to the action of those objects on the sense organs. The classic form of this general view was the representative realism (also called the representative or causal theory) of René Descartes and John Locke, which is still maintained in principle by many scientists. From Berkeley on it suffered much criticism, and its defects led to its being unpopular among philosophers. Modern attempts have been made, however, to remedy these defects and to propose an acceptable theory. The resultant position we shall discuss as critical realism. Although they start from an analysis of perceptual experience and do not argue from the causal processes underlying it, supporters of the sense-datum analysis who are not phenomenalists are forced into one of these kinds of dualist realism.

**REPRESENTATIVE REALISM.** In what is loosely called “seeing a table,” light rays reflected from the table strike the eye, cause chemical changes in the retina, and send a train of impulses along the optic nerve to the brain. The resultant brain activity is then said to cause the mind of the percipient to be directly aware of private *sensa* (Locke called them “ideas”) that represent the shape, color, and other visual properties of the table. A similar account is given for the other senses. The essential point is that perceiving proper is the direct awareness of *sensa*; perceiving

external objects is redefined as perceiving *sensa* caused by them, and so all our awareness is strictly limited to *sensa*. “Represent” is usually interpreted in accordance with the doctrine of primary and secondary qualities—that is, the *sensa* resemble the object in spatiotemporal properties but not insofar as colors, sounds, smells, and other secondary qualities are concerned. Modern analogies of “representing” are the relation between a map or radar screen and the region they cover or between television or movies and the studio events reproduced.

*Merits of representative realism.* Representative realism has important merits. It is the easiest inference from the scientific account of the causal processes up to the brain in all perceiving and fits other scientific evidence. Thus, color blindness and deafness are the result of defects in the sense organs that so affect all subsequent stages in the causal transmission that the resultant *sensa* are different from normal. That electrical stimulation of the brain causes sensations of color, smell, and so on, according to location, seems to confirm the theory, and it can easily accommodate the time lag in perception. Further, by holding that representation does not amount to resemblance in the case of secondary qualities, it can be made to fit the distinction between the world as we see it (that is, the *sensa* grouped as ostensible objects) and the scientific account of material objects, which is in terms of colorless, tasteless, and smell-less elementary particles.

Representative realism also accounts for illusions, dreams, images, hallucinations, and the relativity of perception. Relativity and many illusions result from changes in the stimulation of the sense organs because of distance, medium, angle of sight, and other relevant factors; such changes affect all that follows and so vary the *sensa* caused. Other illusions are the result of misinterpretation of *sensa*. In imagery and dreams the brain activity that occurred in corresponding perceptions is reactivated as the result of internal causes and so brings about the recurrence of similar *sensa*. (The reactivation may be only partial, and the resultant data may be consciously or unconsciously altered by the mind.) Hallucinations are also imagery. Since the images are of a similar character to normally perceived data and are the result of a similar immediate cause in the brain, it is easy to see how they may merge in integrated or triggered hallucinations or how perception may be imaginatively supplemented. The standard explanation of phantom limbs—that they are sensations caused by irritation at the stump of nerves normally coming from the amputated limb—is also accommodated. As perception is confined

strictly to the effects of the causal chain, interference with it en route may readily deceive us.

Finally, representative realism has also traditionally been part of the widely accepted interactionist or dualist account of the relation of mind and body: The body affects mind in perception, mind affects the body in voluntary action. Not all who accept that theory realize that they are saddled with representative realism.

*Defects of representative realism.* Despite its merits, representative realism has some serious defects. If, as it claims, our perceiving is strictly awareness of the mental ideas or *sensa*, it is difficult to see how we can break out of the circle of *sensa* and observe external objects. How can we tell what these objects are like; indeed, how do we know that there are such objects? If we try to verify the existence of the table by touching it, we simply obtain more *sensa*—tactile ones—and if we see our hands touching the table, we are just having visual *sensa*. Whenever we try to peer over the barrier of *sensa*, we just get more *sensa*. This difficulty undermines the analogies used in the theory. Representation is conceived of as something like mapping or photographing, but we know a map represents or a photograph resembles an object because we can observe both and compare them; ex hypothesi, however, we can never strictly observe both objects and *sensa* to compare them. Observing objects is just observing *sensa*, so we do not know that objects and *sensa* resemble each other in primary but not in secondary qualities.

It is often said that representative realism not only leads to skepticism but is also self-refuting, cutting off the branch on which it sits. Its premises and evidence assume that we discover the action of the objects on the sense organs by observing them. Its conclusion—all our perception is of *sensa*—denies that we can do this. However, there would be self-refutation only if the conclusion contradicted the premises, which it need not do if carefully stated. The theory may be regarded as really distinguishing two types of perceiving: perception in its everyday meaning, which is discovering about external objects by means of the senses, and perception proper—direct awareness of *sensa*. It is saying that the first type really amounts to or, better, is really effected by the second type. Thus, granted that by perceiving *sensa* we do discover the nature of objects (at least insofar as their primary qualities are concerned) and their interaction, the first type of perception and the evidence it gives still hold good, and there is no self-refutation. Nevertheless, the skepticism remains, for since our direct awareness is limited to *sensa*,

we do not know that there are objects or what they are like; we only suppose or guess that and what they are.

Even though representative realism need not be self-refuting, it is open to the charge of circularity if considered as an attempt to explain perceiving. It appears simply to transfer perceiving as ordinarily conceived (a face-to-face confrontation) from outside to inside the person; perceiving external objects is now put forward as perceiving private replicas of them, for we look at maps and television pictures in the same way that we look at the countryside. Even if we say perceiving objects is achieved by perceiving *sensa*, there is the same duplication of perceiving, which is thus explained in terms of itself.

Representative realism's view of the mind is rather crude, for it tends to speak almost as if the self or mind were a little person in the head looking at pictures of the outside world. It is not clear how *sensa* can exist in an unextended mind, since they apparently possess shape and size; nor is any serious attempt made to fit the psychological processes of perception into the general scheme.

There are special difficulties for those versions of the theory that claim that in perceiving objects we infer the existence or nature of external objects from our *sensa*. Apart from the inevitable dubiety of such inference, the main objection is that we are never conscious of these inferences nor are we aware of *sensa* as such—that is, as private mental data. If we were, it is difficult to see how the notion of publicly observable causes would occur to us. But the representative theory may simply say that the *sensa* seem to be external (or externally caused) from the start and that any inference is justificatory to deal with skeptics. (This seems to have been Locke's view in his *Essay concerning Human Understanding*, Bk. IV, Ch. xi, Sec. 2.)

**CRITICAL REALISM.** Critical realism is the name primarily given to the views expressed by the American authors of *Essays in Critical Realism*—namely, that the data in perception (that is, what is intuited, what we are directly aware of) are not actually part of external objects but are “character-complexes ... irresistibly *taken*, in the moment of perception, to be the characters of existing outer objects” (p. 20). In veridical perception these characters are the characters of external objects; in illusions they are not. The authors were unfortunately divided over the nature of this datum or character complex, Durant Drake, A. K. Rogers, George Santayana, and C. A. Strong claiming that it was not a mental existent or any kind of existent, but only an essence, a mere logical entity or uni-

versal, whereas A. O. Lovejoy, J. B. Pratt, and R. W. Sellars held that it was a mental existent, a content of sensory experience. It is difficult to grasp what the datum can be if it is not a mental content or existent, and so the second version is the more plausible and is adopted here. Although clearly dualist, it should not be confused with representative realism; in fact, it provides remedies for representative realism's main faults.

The critical realists held that the root of the troubles of representative realism lay in its failure to analyze perceiving or perceptual knowledge. Accepting the ordinary notion of perceiving as intuiting, which means a direct awareness or confrontation, and finding that because of the causal processes and of illusions such awareness was not of external objects, Locke concluded that it must be of intramental ideas and so imprisoned us in the circle of such ideas. The more reasonable conclusion, however, would be that this ordinary notion of perceiving is wrong and that a more careful analysis is needed. This will show that an essential feature of perceiving, even as ordinarily understood, is that it is the way we discover the existence and nature of external objects—that it is, in fact, a claim, often justified, to knowledge. If we appreciate this from the start, we shall not be tempted by the apparently intuitive character of perceiving into an analysis that limits it to ideas, and if we remember that this knowledge claim is not always justified—that is, that there are illusions and errors—we shall avoid the other pitfall of direct realism, in which error becomes inexplicable.

The next step is to realize that though it involves an intuition or direct awareness, perceiving is much more than this. It also involves an active external reference, as is implied by the knowledge claim; we refer this intuited mental content or character complex to an external object—that is, we explicitly judge that it is, or is the character of, an external object or we unreflectingly take it to be this or we immediately react to it as if it were an external object. These modes of reference are differently stressed by different writers, but the point seems to be that they occur in varying degrees according to circumstances. Our perception is sometimes an explicit identification or judgment, or at least it immediately issues in one—for example, we say, “Here’s our bus” or “There’s Tommy”; more often we just see that it is Tommy without formulating any judgment, or our perception that it is our bus and our starting to go and catch it seem indistinguishable, for the reference to the external object is manifest in an immediate physical response.

All the same, in contrast to the behaviorists, the critical realists stressed that there was an intuited mental

content, the character complex of which we were directly aware. Attempts were made to fit the analysis in with current psychology by explaining how this external reference arose in childhood—the apparent externality of the content was with us from the beginning of perceptual discrimination, largely because the external reference was founded in physical response to the object.

There is some similarity between this “reference of an intuited datum to an external object” and the “taking for granted that a sense datum belongs to a material object” of Price’s sense-datum theory, especially since both stress that no distinction between datum and object is drawn by the percipient at the time. But there is a difference in starting point and emphasis. Price began with sense data, treating them as distinct existents and willing to allow that material objects consisted of them. This branch of critical realism began with knowledge of external objects, but, being mental, the content or datum distinguished within it was not regarded as capable of distinct existence and was very difficult—much more so than Price thought—to isolate even subsequently from the associated reference. Also, reference covered a wider set of activities than taking for granted, for it also involved the bodily reactions. In order to stress the relative subordination of the datum, some critical realists spoke of perceiving external objects by means of, guided by, or mediated by, the datum.

Since critical realism can agree that the datum is generated, it is free from the difficulties of the selective theory and can share in the advantages of representative realism. In this version it seems able to avoid the latter’s worst faults. There is no self-refutation, for from the start perceiving is always perception of external objects by means of the intuited data, an analysis that does not deny that we perceive such objects. There is no duplication or circularity, for the direct awareness of the datum is not a replica of perceiving; insofar as it can be distinguished at all, it is much less complex than perceiving, for it involves no identification with external objects and is not in itself directed on them—hence, the map and movie analogies are essentially faulty. Common sense is not being offered an explanation of perceiving in terms of perceiving; it is being shown that perceiving is far more complex than common sense supposes, involving not only causal processes that bring about the datum or mental content but also the psychological processes of reference or response.

Moreover, there need be no skepticism. True, in perceiving we only take the datum to be an external object or its properties, and this may, of course, be erroneous. In a

sense it is always erroneous in that the datum or content is never the object, but normally the taking or reference is correct to the extent that we are perceiving an external object and that the intuited characters also do characterize the external object insofar as primary qualities are concerned; to that extent we are perceiving actual properties or at least projections of them. In general, the claim that perceiving is thus far veridical and amounts to knowledge is said to be the best hypothesis to explain the order and nature of our sense experiences. The realist claim is simply that once ordinary errors and illusions are ruled out by comparing the evidence of different senses or of different persons, the simplest explanation of the situation is that there are external objects causing the sense data or contents and corresponding to them in primary qualities. And this is plausible because if we dismiss as incredible solipsism the view that only oneself and one's own sense experiences exist, then the only real alternative is phenomenalism, a view that has fatal weaknesses and really amounts to proposing a series of deceptive coincidences.

Critical realism is not fully satisfactory, however, particularly if regarded as a theory of perceptual consciousness—that is, as an account of the mental activity that goes on in perception. Thus, the alleged datum or character complex suggests a group of sense data and invites the objections discussed under the entry *Sensa*. A closer examination is required not only of the concepts of datum and reference but also of the general relation of mind and body presupposed in perception and of the nature of mental contents; above all, the theory must take full account of the numerous quasi-interpretative activities that modern psychology has found to be involved in perception.

**See also** Illusions; *Sensa*.

## Bibliography

### GENERAL

A clear, simple introduction to the philosophical problems of perception is Bertrand Russell, *Problems of Philosophy* (London: Williams and Norgate, 1912), and a fuller one is W. P. Montague, *The Ways of Knowing* (New York: Macmillan, 1928). A good account of modern positions on perception is given by T. E. Hill, *Contemporary Theories of Knowledge* (New York: Ronald Press, 1961). Detailed summaries of many realist works are given by W. H. Werkmeister, *A History of Philosophical Ideas in America* (New York: Ronald Press, 1949). Less detailed but more intelligible is John Passmore, *A Hundred Years of Philosophy* (London: Duckworth, 1957); also see Rudolf Metz, *A Hundred Years of British Philosophy* (New York: Macmillan,

1938). Many of the works listed below deal with the topics of more than one section.

### THE REALIST POLEMIC AND NEW REALISM

The main source for new realism is E. B. Holt and others, *The New Realism* (New York: Macmillan, 1912), but new realism owed much to William James; see the papers (dating from 1904) collected in his *Essays in Radical Empiricism* (London and New York: Longman, 1912). For a useful collection of articles from this early period see Roderick M. Chisholm, *Realism and the Background of Phenomenology* (Glencoe, IL: Free Press, 1960). Among other important and often closely reasoned articles are R. B. Perry, "The Ego-centric Predicament," *Journal of Philosophy* 7 (1) (1910): 5–14; Bertrand Russell, "On the Nature of Truth," *PAS* 7 (1906–1907): 28–49; and G. E. Moore, "The Refutation of Idealism," in his *Philosophical Studies* (London: Routledge, 1922), to which compare W. T. Stace's counterattack, "The Refutation of Realism," *Mind* 43 (170) (1934): 145–155. For a general summing up in favor of realism, see D. C. Williams, "The A Priori Argument for Subjectivism," *Monist* 43 (1933): 173–202, "The Inductive Argument for Subjectivism" and "The Inductive Argument for Realism," *Monist* 44 (1934): 80–107, 186–209. For a once influential direct-realism treatment of perceptual problems, see T. P. Nunn, "Are Secondary Qualities Independent of Perception?," *PAS* 10 (1909–1910): 191–218. Ludwig Wittgenstein's language arguments are in his *Philosophical Investigations* (Oxford: Blackwell, 1958), Secs. 256ff.; for criticisms see Carl Wellman, "Wittgenstein and the Ego-centric Predicament," *Mind* 68 (270) (1959): 223–233, or the symposium "Can There Be a Private Language?," *PAS*, Supp. 28 (1954).

### PERSPECTIVE REALISM AND ALLIED THEORIES

Perspective realism and allied theories are stated by Evander Bradley McGilvary in *Toward a Perspective Realism* (La Salle, IL: Open Court, 1956) and in "Perceptual and Memory Perspectives," *Journal of Philosophy* 30 (1933): 310ff. Older versions are by Samuel Alexander, "On Sensations and Images," *PAS* 10 (1909–1910): 1–35, and H. A. Prichard, *Kant's Theory of Knowledge* (Oxford: Clarendon Press, 1909), Ch. 4. Despite its title, G. Dawes Hicks, *Critical Realism* (London: Macmillan, 1938), gives a theory of appearing. Such theories are lucidly discussed by Roderick M. Chisholm, "The Theory of Appearing," in *Philosophical Analysis*, edited by Max Black (Ithaca, NY: Cornell University Press, 1950). C. D. Broad, *The Mind and Its Place in Nature* (London: Kegan Paul, 1925), and H. H. Price, *Perception* (London: Methuen, 1932), criticize these theories carefully from a sense-datum standpoint, though Price's article "Illusions," in *Contemporary British Philosophy*, edited by H. D. Lewis (London, 1956), Vol. 3, defends a limited perspective realism.

### COMMONSENSE REALISM

For commonsense realism see G. E. Moore's "A Defense of Common Sense" and "Proof of an External World," which are in his *Philosophical Papers* (London: Allen and Unwin, 1959), but also see his *Some Main Problems of Philosophy* (London: Allen and Unwin, 1953). The staunchest more recent defender of common sense against the argument from illusion is J. L. Austin in his lucid and lively *Sense and*



*Sensibilia* (London: Oxford University Press, 1962), and somewhat similar views are clearly and concisely expressed by Anthony M. Quinton, "The Problem of Perception," *Mind* 64 (253) (1955): 28–51. Gilbert Ryle's "Sensations," in *Contemporary British Philosophy*, edited by H. D. Lewis (London, 1956), Vol. III, and Ryle's *Dilemmas* (Cambridge, U.K.: Cambridge University Press, 1954) try to deal also with the causal argument in a nontechnical manner. D. M. Armstrong, *Perception and the Physical World* (London: Routledge and Paul, 1961), defends direct realism but in so doing is driven toward behaviorism.

#### REPRESENTATIVE REALISM (OR THE CAUSAL THEORY)

For early statements of representative realism, see René Descartes, *Principles of Philosophy*, Pt. IV, and John Locke, *Essay concerning Human Understanding*. Bk. 2, Ch. 8. Representative realism is assumed by many modern neurologists, though often not under its philosophical title. Walter Russell Brain states and discusses it as "physiological idealism" in "The Neurological Approach to the Problem of Perception," *Philosophy* 21 (79) (1946): 133–146, reprinted with further consideration of perception in his *Mind, Perception and Science* (Oxford: Blackwell, 1951); he gives a further defense of his position in his *The Nature of Experience* (London: Oxford University Press, 1959). J. C. Eccles, *The Neurophysiological Basis of Mind* (Oxford: Clarendon Press, 1953), pp. 279–281, outlines the theory as if it were fact. J. R. Smythies, *Analysis of Perception* (London: Routledge and Paul, 1956), puts forward an improved form of it closer to critical realism.

#### CRITICAL REALISM

The main source for critical realism is Durant Drake et al., *Essays in Critical Realism* (London: Macmillan, 1920), which reveals the differences as well as agreements; see also other works by the essayists, especially R. W. Sellars's comprehensive *The Philosophy of Physical Realism* (New York: Macmillan, 1932), his general apologia, "A Statement of Critical Realism," *Revue internationale de philosophie* 1 (1938–1939): 472–498, and A. O. Lovejoy's impressive general defense, *The Revolt against Dualism* (La Salle, IL: Open Court, 1930). R. J. Hirst, *The Problems of Perception* (London, 1959), also discusses commonsense and representative realism and reaches a somewhat similar position. A primarily pragmatist view that has affinities to and criticism of critical realism is C. I. Lewis, *Mind and the World Order* (New York: Scribners, 1929).

#### OTHER RECOMMENDED TITLES

Almeder, Robert. *Blind Realism*. Lanham, MD: Rowman & Littlefield, 1992.  
 Alston, William. *A Realist Conception of Truth*. Ithaca, NY: Cornell University Press, 1996.  
 Benacerraf, P. "Mathematical Truth." *Journal of Philosophy* 70 (1973): 661–680.  
 Brink, D. *Moral Realism and the Foundations of Ethics*. Cambridge, U.K.: Cambridge University Press, 1989.  
 Carnap, Rudolf. "Empiricism, Semantics, and Ontology." In his *Meaning and Necessity*. 2nd ed. Chicago: University of Chicago Press, 1956.  
 Churchland, Paul M. *Scientific Realism and the Plasticity of Mind*. Cambridge, U.K.: Cambridge University Press, 1979.

Devitt, Michael. *Realism and Truth*. Oxford: Blackwell, 1984.  
 Dummett, M. A. E. "Realism." In *Truth and Other Enigmas*. London: Duckworth, 1963.  
 Goodman, Nelson. *Ways of Worldmaking*. Indianapolis: Hackett, 1978.  
 Moser, Paul K. *Philosophy after Objectivity*. New York: Oxford University Press, 1993.  
 Putnam, Hilary. *The Many Faces of Realism*. La Salle, IL: Open Court, 1987.  
 Quine, W. V. O. *Word and Object*. Cambridge, MA: MIT Press, 1960.  
 Quinton, Anthony. *The Nature of Things*. London: Routledge and Kegan Paul, 1977.  
 Searle, John R. *The Construction of Social Reality*. New York: Free Press, 1995.  
 Sellars, Wilfrid. *Science, Perception, and Reality*. Atascadero, CA: Ridgeview Press, 1991.  
 Smart, J. J. C. *Philosophy and Scientific Realism*. London: Routledge and Kegan Paul, 1963.  
 van Fraassen, Bas. *The Scientific Image*. Oxford: Clarendon Press, 1980.  
 Wright, C. J. G. *Realism, Meaning, and Truth*. Oxford: Blackwell, 1987.  
 Wright, Crispin. *Truth and Objectivity*. Cambridge, MA: Harvard University Press, 1992.

R. J. Hirst (1967)

Bibliography updated by Benjamin Fiedor (2005)

## REALISM [ADDENDUM]

Contemporary philosophical "realism" is not a single thesis but rather a diverse family of positions, unified chiefly by their invocation of certain characteristic images and metaphors. The realist about a region of discourse typically holds, for example, that our central commitments in the area describe a world that exists anyway, independently of us; that cognition in the area is a matter of detection rather than projection or constitution; and that the objects of the discourse are real things and not just linguistic or social constructions. Debates over realism defined in terms such as these persist in nearly every philosophical subdiscipline: from ethics and the philosophy of mind to the philosophy of science and the philosophy of mathematics. (Although it is common to describe a philosopher as a realist or nonrealist *tout court*, realism in one area is generally independent of realism in another, and advocates of global realism and its opposite number, global nonrealism, are comparatively rare.) Contemporary discussion is concerned in part with the evaluation of these discipline-specific realist theses. But it is also concerned (and increasingly so) with the more basic question of how exactly the realist's distinctive imagery is to be understood.

We may epitomize the realist's stance by saying that to be a realist about a region of discourse is to regard it as describing a genuine domain of objective fact. But what is it for a discourse to describe a "domain of fact"? And what is it for a domain of fact to be "objective"? These questions are usefully approached by attempting a taxonomy of the alternatives to realism. The nonrealist rejects the realist's rhetoric of objectivity. But this rejection can take a number of more determinate forms, and their variety sheds considerable light on what realism requires.

The realist's most basic commitment is to the view that statements in the target area purport to describe a world—to say how things stand with some distinctive range of objects or facts. This claim is often glossed as the minimal requirement that statements in the area be capable of truth or falsity. Realism is thus opposed at this most basic level to nonfactualism (also called irrealism or noncognitivism): the view that declarative statements in the target area cannot be evaluated as true or false and so cannot serve a descriptive function. Nonfactualist theses have been advanced mainly in moral philosophy, where it has been suggested that moral utterances serve to express emotional attitudes (emotivism: Blackburn, 1984; Gibbard, 1990) or to endorse or proscribe certain courses of action (prescriptivism: Hare, 1963; cf. Geach, 1963). But they have occasionally been proposed in other areas. Formalism in the philosophy of mathematics (the view that mathematics is a game with meaningless marks, manipulated according to formal rules) and instrumentalism in the philosophy of science (according to which theoretical statements function as uninterpreted tools for deriving predictions about future experience) are further examples of this kind of nonrealism.

To say that a region of discourse purports to describe a world is to say more than that its central commitments are apt for truth. It is to say, in addition, that they are aimed at truth—that they are typically put forward as genuine assertions about how things stand with their ostensible subject matter. Realism is thus opposed at this second level to fictionalism, the view that seeming assertions in the target area, though capable of truth, are in fact designed only to provide representations that are somehow "good" or "interesting" or "useful" for certain purposes. Fictionalist approaches have been developed mainly in the philosophy of science, where Bas van Fraassen's constructive empiricism provides a useful example (van Fraassen, 1980; cf. Churchland and Hooker, 1985). Van Fraassen agrees with the scientific realist, against the instrumentalist, that theoretical statements possess definite truth conditions and so constitute

genuine representations of unobservable structures. However, he further maintains, this time against the realist, that the truth-value of a theory is irrelevant to its acceptability from the standpoint of science. The aim of science on van Fraassen's view is empirical adequacy: the correct description of the observable world. Theories may posit unobservable things. But a good scientific theory—one that satisfies to some high degree all of the aspirations implicit in the scientific enterprise—may be largely false in its account of such matters, so long as it is a reliable guide to the observable world. In advancing a theory in what seems to be the assertoric mode, the scientist shows only that he accepts it as empirically adequate. Van Fraassen's fictionalism thus consists centrally in his contention that the endorsement of a scientific theory does not involve the belief that it is true or that the unobservables it posits exist. Generalizing, we may say that realism involves, in addition to the semantic thesis of truth aptitude, the pragmatic thesis of truth directedness, according to which the target discourse aims at truth, and the endorsement of a claim is normally an expression of one's belief that it is true. (See Field, 1980, for a fictionalist approach to the philosophy of mathematics.)

Before we have a position that is recognizable as realist we must add one further ingredient. It is not enough that our central commitments aspire to truth. They must also be true, or at least not wildly mistaken. Realism is thus opposed at this third level to a conception of the target area as involving a fundamental mistake about what the world contains. This "error-theoretic" alternative to realism is typified by J. L. Mackie's view of morality (Mackie, 1977). According to Mackie, ethical discourse purports to describe a range of objective prescriptions, constraints on action that are somehow built into the fabric of nature. But since it can be shown (Mackie held) that there are no such items, it follows that morality is based on a mistake—the entities it purports to describe do not exist; the properties it trades in are not instantiated—and hence that moral discourse demands reconstrual, if not outright rejection. A more familiar instance of the error-theoretic approach is atheism, the view that theological discourse is vitiated by the mistaken supposition that God exists. Agnostic versions are also possible, though in fact they have played no significant role outside the philosophy of religion.

A philosopher who holds that our core commitments in an area succeed in providing a true account of their intended subject matter may be called a minimal realist about that area. It is sometimes suggested that there is nothing more to realism than this minimal view

and hence that once the questions of truth aptitude, truth directedness, and truth have been settled, there is no further space for debate about whether the discourse is to be understood “in a realistic fashion.” There are, however, at least two reasons to resist this claim.

The first concerns the classification of reductionist positions. The behaviorist thesis that psychological statements can be reduced without remainder to claims about overt bodily movements and the like is clearly compatible with minimal realism about the mind. And yet the view that there is nothing more to being in pain than exhibiting “pain behavior” has generally been regarded as a clear alternative to a robust realism about mental states. It has thus become customary to insist that the realist’s commitment to the truth of our views in the target area be a commitment to their truth on a literal or face-value construal (Blackburn, 1984, chap. 5; cf. Wright, 1983). The behaviorist translation of a simple psychological statement such as “Nadja is dreaming of Paris” will typically be a long conjunction of conditional claims describing the outward behavior Nadja would exhibit if prompted by various stimuli. But this paraphrase has a very different “surface form” from the psychological claim whose meaning it is meant to capture. And this suggests that on the behaviorist’s account, the correct interpretation of psychological statements is not a face-value interpretation and hence that while he may endorse a version of minimal realism about the mental, the behaviorist should not be classed as a realist without qualification.

The second and more serious reason to resist the identification of realism with minimal realism is that minimal realism by itself involves no commitment to the mind independence or objectivity of the disputed subject matter. Immanuel Kant’s transcendental idealism has generally been regarded as a paradigmatic alternative to full-blown realism about the external world; and yet it is fully compatible with minimal realism as defined above. Objects in space and time are real, for Kant, in the sense that much of what common sense and science have to say about them is literally true. And yet there is another sense in which they are not fully real. The structure of the spatiotemporal world is “conditioned” for Kant by the structure of the mind that experiences it. Empirical investigation is therefore not addressed to a domain of fact that is altogether “independent of us.” Clearly, Kant’s position should not be described as a species of realism without qualification, its consistency with minimal realism notwithstanding.

Much of the most important work on realism has been devoted to explicating the commitment to objectiv-

ity that seems a necessary component of any fully realist position. The most natural thought is to identify objectivity directly with a straightforward sort of mind independence. A state of affairs will then count as objective if it would have obtained (or could have obtained) even if there were no minds or mental activity. But this precludes realism about the mind itself and also about any discourse in the social sciences that concerns itself with the products of human thought and action. And this is implausible. It should be possible to be a realist about psychology, for example, while conceding that the facts it describes are obviously mind dependent in the sense that they would not have obtained if there were no minds.

One influential approach to this problem is due to Michael Dummett, whose work is largely responsible for the current prominence of realism as a theme in Anglophone philosophy (Dummett, 1978; cf. Wright, 1992). On Dummett’s view the dispute over realism, though ultimately an issue in metaphysics, is best approached by recasting it as a dispute within the philosophy of language about how to construct a theory of meaning for the target discourse. A theory of meaning in Dummett’s sense is a representation in propositional form of what a competent speaker knows in virtue of which he understands his language. Dummett identifies realism with the view that a meaning theory must take the form of a classical two-valued semantic theory: an assignment of truth conditions to sentences that respects the principle of bivalence, according to which every sentence is determinately either true or false. Realism’s slogan is: To understand a sentence is to know its truth condition. The leading alternative—sometimes called semantic antirealism—holds instead that to understand a sentence is to know the conditions under which it is correctly asserted. A view of this sort assigns each declarative sentence a class of “verification conditions,” each of which must be the sort of condition a competent human being can in principle recognize as obtaining. A semantic theory constructed upon such a basis will generally fail to respect bivalence. The only notion of truth it makes available will be epistemically constrained: Truth will be identified with knowable truth, and falsity with knowable falsity. On a view of this sort we shall not be entitled to say in advance that every well-formed question must have an answer, or that every statement of the form “*p* or not-*p*” must be true. This rejection of bivalence (and the closely related law of excluded middle) is the hallmark of semantic antirealism. To suppose that the only notion of truth we possess for a region of discourse is an epistemically constrained one is to suppose that the facts in the area are (as it were) cut to fit our intellectual capacities. Conversely, to insist that bivalence

must hold regardless of our cognitive limitations is to conceive of the facts at which our thought is directed as obtaining (in one sense) independently of us.

A closely related proposal has been advanced by Hilary Putnam, (1978, 1987). Putnam identifies full-blown “metaphysical” realism directly with the view that truth is epistemically unconstrained. As Putnam frames the issue, the metaphysical realist’s characteristic thought is that an ideal theory might be false, where an ideal theory is one that satisfies perfectly every criterion we normally employ in deciding what to believe in the target area. In the scientific case, for example, an ideal theory would be one that supplies accurate predictions of experimental outcomes while simultaneously displaying every internal theoretical virtue that scientists consider in the context of theory choice: simplicity, elegance, explanatory power, “intrinsic plausibility,” and the like. It is natural to suppose that such a theory could be false. After all, the theoretical virtues that provide our only grounds for choice among empirically equivalent hypotheses seem importantly subjective. A theory that strikes us as particularly powerful because it provides informative answers to interesting questions might strike creatures with different interests as unacceptably silent on important matters; a theory that strikes us as “intrinsically plausible” might strike creatures with different histories or cultures as strange and unlikely. The thought that an ideal theory can be false thus seems a natural expression of an appropriate human modesty, according to which we can have no guarantee in advance that our contingent, biologically, and historically conditioned sense of theoretical virtue must be a reliable guide to the facts about the physical world.

Putnam rejects this natural thought. Metaphysical realism presupposes a concept of truth that is radically divorced from our notion of correct assertion. But according to Putnam such a concept is unattainable. Putnam’s case for his view, like Dummett’s, defies simple summary; but in rough outline it proceeds as follows: The only serious effort to explain an epistemically unconstrained notion of truth is a version of the correspondence theory of truth. This approach proceeds in two stages. First, subsentential expressions such as names and predicates are associated with objects and properties as their referents. Then truth as a feature of sentences is defined recursively according to a scheme well known to logicians. Putnam’s central contention is that there is no credible account of the first stage. Every attempt to explain in realist terms how a word manages to refer to one object rather than another—that is, every attempt to

explain how language “hooks on” to the world—is either plainly unsatisfactory or implies a radical indeterminacy of reference.

Putnam’s alternative is to identify truth directly with “ideal acceptability,” a position he calls internal realism. The position is realist, not simply because it is compatible with what has here been called minimal realism, but also because it eschews reductionism while remaining compatible with all of the ordinary denials of mind dependence that are part of our scientifically informed worldview. Since it is plainly correct by ordinary standards to assert that mountains exist even when no one is aware of them, the internal realist will agree that mountains do not depend in this literal sense on our thought and are therefore in that sense objective. Still, the view does imply an internal connection at the global level between the way the world is and the way we are disposed to conceive of the world in what Charles Sanders Peirce called “the ideal limit of inquiry.” According to the internal realist, we should not say (as the idealist would) that the mind somehow constructs the world but rather that “the mind and the world together make up the mind and the world” (Putnam, 1978).

It remains uncertain whether the efforts of Dummett, Putnam, and others to describe a plausible alternative to realism on the matter of objectivity can succeed. It is to be noted that the arguments they provide indict any epistemically unconstrained notion of truth whatsoever, and hence that if they succeed at all they imply a global antirealism according to which every region of human thought that satisfies the condition of minimal realism is directed at a region of fact that is somehow constituted in part by our thought about it. But this can be rather hard to believe. The difficulty emerges most dramatically when we consider discourse about the past. Most of us are inclined to believe that every (nonvague) question about the past must have an answer. There is a fact of the matter, we suppose, as to whether Genghis Khan was right-handed, even if we cannot in principle obtain any pertinent evidence. But it is likely that neither “Genghis Khan was right-handed” nor its denial is assertible. Any view according to which this implies that the statement is neither true nor false is therefore bound to strike us as initially incredible. Perhaps more important, there is reason to doubt whether a commitment to an epistemically constrained notion of truth always implies a rejection of the realist’s rhetoric of objectivity and independence. It is conceivable, for example, that a moral realist for whom the demands of morality are entirely independent of our passions and interests might nonetheless insist that the

moral facts—because they represent rationally compelling demands on human action—must be accessible in principle to human beings. Moral truth would then be epistemically constrained; and yet the realist’s rhetoric of objectivity and independence would not be undermined.

To be a realist about a region of discourse is to hold at a minimum that our core commitments in the area are largely true when interpreted “at face value.” However, this minimal characterization fails to capture the realist’s commitment to the objectivity or mind independence of his subject matter. In some cases this further commitment can be understood as the requirement that the concept of truth appropriate to the target area be epistemically unconstrained. It remains unclear, however, whether this characterization is adequate to every case. The search for a fully general account of the realist’s commitment to objectivity is perhaps the central open question in this part of philosophy.

**See also** Atheism; Dummett, Michael Anthony Eardley; Idealism; Kant, Immanuel; Mackie, John Leslie; Meaning; Metaphysics; Noncognitivism; Peirce, Charles Sanders; Philosophy of Language; Philosophy of Mind; Philosophy of Science, History of; Philosophy of Science, Problems of; Putnam, Hilary; Reference; Tarski, Alfred; Truth; Van Fraassen, Bas.

### **Bibliography**

- Blackburn, S. “How to Be an Ethical Anti-Realist.” In *Essays in Quasi-Realism*. New York: Oxford University Press, 1993.
- Blackburn, S. *Spreading the Word*. Oxford: Clarendon Press, 1984.
- Churchland, P., and C. Hooker, eds. *Images of Science*. Chicago: University of Chicago Press, 1985.
- Devitt, M. *Realism and Truth*. Cambridge, MA: Blackwell, 1991.
- Dummett, M. *The Seas of Language*. Oxford: Clarendon Press, 1993. Especially the essays “What Is a Theory of Meaning II” and “Realism.”
- Dummett, M. *Truth and Other Enigmas*. Cambridge, MA: Harvard University Press, 1978. Especially the essays “Realism,” “Truth,” and “The Philosophical Basis of Intuitionistic Logic.”
- Field, H. *Science without Numbers*. Princeton, NJ: Princeton University Press, 1980.
- Geach, P. “Ascriptivism.” *Philosophical Review* 69 (1960): 221–225.
- Gibbard, A. *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press, 1990.
- Hare, R. M. *Freedom and Reason*. Oxford: Clarendon Press, 1963.
- Honderich, T., ed. *Morality and Objectivity: A Tribute to J. L. Mackie*. London: Routledge and Kegan Paul, 1985.

Mackie, J. L. *Ethics: Inventing Right and Wrong*. New York: Penguin, 1977.

Putnam, H. *The Many Faces of Realism*. La Salle, IL: Open Court, 1987.

Putnam, H. *Meaning and the Moral Sciences*. Boston: Routledge and Kegan Paul, 1978.

van Fraassen, B. *The Scientific Image*. Oxford: Clarendon Press, 1980.

Wright, C. *Frege’s Conception of Numbers as Objects*. Aberdeen: Aberdeen University Press, 1983.

Wright, C. *Truth and Objectivity*. Cambridge, MA: Harvard University Press, 1992.

**Gideon Rosen (1996)**

## REALISM, LEGAL

See *Legal Realism*

## REALISM AND NOMINALISM

See *Universals, A Historical Survey*

## REALISM AND NATURALISM, MATHEMATICAL

Many versions of realism in mathematics are intimately related to versions of naturalism. The purpose of this article is to explore relationships between the various views, and, briefly, the main opposition to them. The focus here is exclusively on mathematics. So, for example, “Platonism” is to be read as “Platonism about mathematics.” This entry does not claim to do justice to the subtle and detailed works of everyone who works in the philosophy of mathematics, or even everyone who defends versions of realism and/or naturalism. Instead, this entry seeks to provide a useful road map of an important part of the territory.

In broad terms, *realism* is the view that mathematics is objective: independent of the lives, customs, language, and form of life of mathematicians. This statement is deliberately indeterminate. What aspects of mathematics are being discussed? What, exactly, is it independent of? And what is it to be independent? What is it to be objective? In philosophy there is little that one can take for granted.

## REALISM

There are at least two forms of realism: realism in ontology, which concerns mathematical objects, and realism in truth value, which concerns mathematical truth. *Realism in ontology* is the view that mathematical *objects*, such as numbers, sets, functions, and geometric points exist independently of the mathematician. *Prima facie*, these mathematical objects do not occupy physical space; they exist eternally and are not created or destroyed; and they do not enter into causal relationships with either each other or with physical objects. Because Platonic forms share these features, realism in ontology is sometimes called “Platonism” or, as Geoffrey Hellman (1989) dubs it, “objects Platonism.” This sort of Platonism is sometimes written with a lowercase “p,” perhaps to mark some distance from Plato. For the realist in ontology, mathematical propositions are taken at face value, as statements about mathematical objects. The theorem that 101 is a prime number just is the statement that a given object, the number 101, enjoys a certain property, primeness. The sentence “101 is a prime number” has the same logical form as “Socrates is Greek.” Most versions of realism in ontology have it that mathematical truth is necessary, in a deep metaphysical sense: If the subject matter of mathematics is as these realists say it is, then typical propositions about mathematical objects—the principles of pure mathematics, for example—do not suffer from the contingencies of science or ordinary statements about ordinary physical objects.

Probably the most difficult problems associated with realism in ontology are in epistemology (see Benacerraf 1973). The realist declares that mathematics is about a realm of *prima facie* abstract, causally inert, and eternally existing objects. How can human beings ever come to know anything about these objects? How can humans have reliable, justified beliefs about such objects? The way people come to know things about physical objects typically involves some sort of causal contact between people, the knowers, and the objects (e.g., seeing them). This is ruled out with mathematical objects. Presumably, most of the beliefs that mathematicians have about mathematical objects are true. Mathematicians are reliable indicators of how things are in the mathematical realm. How does one explain this reliability (see Field 1989, essay 7)?

One resolution to these problems is to postulate a special faculty that humans have, an intuition, that links humans to the mathematical realm. Such was Plato’s own solution to the analogous problem concerning Forms. Some of the logician Kurt Gödel’s (1944, 1964) remarks can be interpreted along these lines:

Despite their remoteness from sense experience, we do have something like a perception also of the objects of set theory, as is seen from the fact that axioms force themselves upon us as being true. I don’t see any reason why we should have less confidence in this kind of perception, i.e., in mathematical intuition, than in sense perception ... It should be noted that mathematical intuition need not be conceived of as a faculty giving an *immediate* knowledge of the objects concerned. Rather, it seems that, as in the case of physical experience, we *form* our ideas also of those objects on the basis of something else which is immediately given ... It by no means follows ... that the data of this ... kind, because they cannot be associated with actions of certain things upon our sense organs, are something purely subjective ... Rather they ... may represent an aspect of objective reality, but, as opposed to the sensations, their presence in us may be due to another kind of relationship between ourselves and reality. (Gödel 1964, p. 484)

A philosopher who is inclined this way has the task of trying to square the presence of mathematical intuition with the current scientific view of a human being as a thoroughly physical organism in a physical universe.

**LOGICISM.** Another strategy for epistemology comes from *logicism*, the view that mathematical truth is a species of logical truth. The epistemology for mathematics is thus the epistemology for logic. The most detailed developments are those of Gottlob Frege (1884, 1893) and Alfred North Whitehead and Bertrand Russell (1910). Of those, Frege was a realist in ontology, at least for arithmetic and analysis. So for Frege, logic has an ontology—there are “logical objects.” Numbers are constructed out of logical objects.

In attempting to define the natural numbers and the general notion of *natural number*, Frege (1884, §63) proposed the following principle, which has become known as “Hume’s principle”:

For any concepts  $F$ ,  $G$ , the number of  $F$ ’s is identical to the number of  $G$ ’s if and only if  $F$  and  $G$  are equinumerous.

Two concepts are equinumerous if they can be put in one-to-one correspondence. Frege showed how to define equinumerosity without invoking natural numbers. In the end, he balked at taking Hume’s principle as the ulti-

mate foundation for arithmetic, and went on to provide an explicit definition of the natural numbers in terms of concepts and their extensions. The number two, for example, is the extension (or collection) of all concepts that hold of exactly two elements. Unfortunately, the inconsistency in Frege's theory of extensions, as shown by Russell's paradox, marked a tragic end to Frege's logicist program.

Variations of Frege's approach are vigorously pursued in the early twenty-first century, in the work of Crispin Wright, beginning with (1983), and others such as Bob Hale (1987) and Neil Tennant (1997). The idea is to bypass the treatment of extensions and to work with Hume's principle, or something like it, directly. On this *neo-Fregean* approach, Hume's principle is taken to be an explanation of the concept of "number." It is an implicit definition, true by stipulation. Frege's own technical development shows that the Peano postulates can be derived from Hume's principle in a standard, higher-order logic. Indeed, the only essential use that Frege made of extensions was to derive Hume's principle—everything else concerning numbers follows from that.

**HYPOTHETICAL-DEDUCTIVE APPROACH.** Another popular strategy for epistemology comes from an overarching hypothetical-deductive approach. The argument begins with the observation that virtually all of science is formulated in mathematical terms. One cannot believe in the truth of physics, say, without also accepting the mathematics that occurs in it. Thus, mathematics is confirmed to the extent that science is. In short, because mathematics is indispensable for science, and because science is well-confirmed and (approximately) true, one can conclude that mathematics is well-confirmed and true as well. This "indispensability argument" is attributed to W. W. O. Quine; a clear articulation is found in Hilary Putnam's *Philosophy of Logic* (1971, ch. 5) (see also Colyvan 2001).

**STRUCTURALISM.** According to *structuralism*, the subject matter of arithmetic, for example, is the pattern common to any infinite system of objects that has a distinguished initial object, which plays the role of zero, and a successor relation or operation that satisfies the induction principle. The arabic numerals exemplify this *natural number structure*, as does an infinite sequence of distinct moments of time, an infinite sequence of discrete points in space, and so on. Similarly, real analysis is about the real number structure, set theory is about the set-theoretic-hierarchy structure, and topology is about topological structures. According to the *ante rem* version

of this view, the natural number structure, for example, exists independently of whether it has instances in the physical world, or any other world for that matter (see Shapiro 1997, Resnik 1997, also Parsons 1990). This is an ontological realism. The number six, for example, is a place in the natural number structure, the seventh place (if one begins with zero). Because, on the view in question, the structure exists objectively, then so do its places. Structuralists have proposed various epistemological strategies, ranging from pattern recognition, linguistic abstraction, implicit definition (much like neo-logicism), and postulation via indispensability (with the Quinean). One line, shared with the full-blooded platonism articulated by Mark Balaguer (1998), holds that the realm of structures is so robust that every coherent axiomatization is true of at least one structure. So the sticky problem concerning knowledge of mathematical objects reduces to knowledge of the coherence of an axiomatization.

**THE OPPOSITION: ANTIREALISM IN ONTOLOGY.** Speaking logically, the opponents of realism in ontology fall into two camps. One group holds that numbers, functions, sets, points, and the like exist, but not objectively. Mathematical objects are not independent of the mind, language, conventions, or the form of life of the mathematician or the mathematical/scientific community. According to traditional intuitionism, for example, mathematical objects are mental constructions (e.g., Brouwer 1912, 1948; Heyting 1956). This is an idealism of sorts. Some intuitionists have explicitly Kantian roots, tying mathematical construction to the forms of pure intuition (typically of time). Another ontological antirealist view sees mathematical objects as social constructions.

The other way to reject ontological realism is to hold that there are no distinctive mathematical objects at all. There simply are no numbers, sets, functions, points, and so on. This is called *nominalism*. Again, it comes in two varieties. On one of them, mathematical assertions keep a straightforward, face-value reading. So the statement that every natural number is prime is vacuously true, because there are no natural numbers. "Seven is prime" is either false or lacks truth-value, depending on how nondenoting singular terms are handled. On this view, mathematical objects are likened to characters and objects in fiction. The sentence that seven is prime is of a piece with "Miss Marple is nosy." Of course, fictionalists do not recommend that mathematicians settle their questions via the literal, face-value reading of their assertions. Either they advert to a "truth in mathematics" akin to "truth in the story" for fiction, or else they provide some other purpose for mathematics beyond seeking mathematical truth (see

Field 1989). Fictionalism is an error-theory about mathematics.

The other variety of nominalism provides alternate, non-face-value readings of mathematics. So the statements of mathematics come out true or false, without presupposing a mathematical ontology. The modal structuralist, for example, reads a statement such as “there are infinitely many prime numbers” as “any exemplification of the natural number structure has infinitely many places, each of which satisfied the property of being prime in that structure.” Charles Chihara (1990, 2004) provides versions of various mathematical theories in terms of possible linguistic constructions. One interesting issue concerns the relationship between the “nominalized” assertions and their original counterparts (see Burgess and Rosen 1997).

**REALISM IN TRUTH-VALUE.** These nominalistic programs lead to another major type of realism concerning mathematics. Georg Kreisel is often cited as suggesting that the important questions in the philosophy of mathematics do not concern the existence of mathematical objects, but rather the objectivity of mathematical assertions. Let us define *realism in truth-value* to be the view that mathematical statements have objective and nonvacuous truth-values independent of the minds, languages, and conventions of mathematicians.

Once again, the opponents to this view logically fall into two categories, depending on what is being denied. The radical opposition holds that mathematical statements have no nonvacuous truth-values at all. The fictionalist, noted above, is the primary and perhaps only occupant of this category. It is difficult to conceive of a projectivism or expressivism concerning mathematics.

The less radical versions of truth-value irrealism allow that mathematical statements have truth-values, but these are not independent of the minds, languages, and conventions of mathematicians. The traditional intuitionists, as described above, fit this bill. Because, for them, mathematical objects are mental constructions, mathematical assertions relate to the activity of construction. Contemporary intuitionists, following Michael Dummett (1977, 1978) also fit this bill, holding that all truths are knowable, on broadly semantic grounds.

Realism in ontology is naturally allied with realism in truth-value. To get from the former to the latter, one just insists that the sentences of mathematics be read literally, at face value. If, for example, “seven” is a genuine singular term, and the sentence “seven is prime” is objectively true, then, it seems, “seven” denotes something, namely, the

number seven. And it exists objectively. Conversely, a realist in ontology gets to realism in truth-value by insisting that the typical propositions concerning the interrelations of the mind-independent mathematical objects are themselves objective.

Nevertheless, the connections between these realisms are not forced by logical connections that are obvious to all. As noted, many nominalists are realists in truth-value. They reject the face-value reading of mathematical assertions. At least one prominent philosopher of mathematics goes in the opposite direction. Neil Tennant (1987, 1997) holds that mathematical objects exist objectively, of necessity, and yet he adopts a Dummettian antirealism concerning truth-value.

## NATURALISM

Unfortunately, the word “naturalism” has become something of a term of art, and it is hard to find a common theme that underlies every view that goes by that name. Perhaps most of them share a certain deference to the natural sciences. Quine characterizes naturalism as “the abandonment of first philosophy” and “the recognition that it is within science itself ... that reality is to be identified and described” (Quine 1981, p. 72; see also 1969). The idea is to see philosophy as continuous with the sciences, not prior to them in any epistemological or foundational sense.

**QUINEAN NATURALISM.** The naturalist accepts the existence of the theoretical entities, such as forces and electrons, that occur in the most up-to-date scientific theories. Current science describes the world in such terms, and it runs against the theme of naturalism to reject them on philosophical grounds, adopting some sort of instrumentalism or constructive empiricism. When it comes to mathematics, however, naturalists differ. As seen with the aforementioned indispensability argument, Quine himself accepts mathematics to the extent—but only to the extent—that it is needed in science. It is impossible to do physics, or just about any other science for that matter, without invoking real analysis. So the theorems of real analysis are confirmed to the extent that the various scientific theories are confirmed, and these theories are the best ones available. So Quine accepts the truth of real analysis. Moreover, some of the traditional, Platonic themes have naturalistic counterparts. For example, the eternity of mathematical objects corresponds to the fact that mathematical assertions are not inflected with tense.

*Naturalized epistemology* is the application of Quinean naturalism to the study of knowledge. The



philosopher sees the human knower as a thoroughly natural being within the physical universe. Any faculty that the philosopher invokes to explain knowledge must involve only natural processes amenable to ordinary scientific scrutiny.

This theme exacerbates the epistemic problems with realism. Platonic apprehension of a detached mathematical universe is ruled out from the start, as a nonnatural process. The challenge to the ontological realist is to show how a physical being in a physical universe can come to know about *abstracta* such as mathematical objects. There may be no refutation of realism in ontology, but there is a deep challenge to it. The advocate of indispensability cites the role of mathematics in science. The idea is that mathematics is known the same way that science is. However, it is not enough to leave it at that. The advocate of realism in ontology should delimit the exact role that mathematics plays in science. How, for example, is it possible for a casually isolated realm of abstracta to shed light on the interactions of physical matter? An answer to this would go a long way toward solving the epistemological puzzles.

Notice that, at best, the indispensability argument delivers the *truth* of the principles of real analysis. If one assumes that science is objective, then there is realism in truth-value. It is not clear that the Quinean naturalist is also committed to realism in ontology, despite Quine's own tendencies in that direction. This depends on whether naturalism requires the philosopher to accept the pronouncements of mathematical science at face value. Quine famously calls for regimentation of ordinary and scientific discourse, to clean up the ontological commitments. One can see some of the aforementioned nominalistic programs in this spirit. Some of them show (or try to show) how mathematics can be true without presupposing the existence of distinctively mathematical objects (Hellman 1989). And this truth is all that is needed in science, or so the argument goes.

Other nominalists take issue with the indispensability argument itself. They show how science could proceed without mathematics, or at least without mathematics as it is standardly understood (Field 1980, Chihara 2004). This is also perhaps in the spirit of naturalism.

Quine's own realism extends to real analysis, functional analysis, and perhaps a bit more. But it stops there. Quine does not accept the truth of the higher reaches of set theory unless and until it finds application in science. In fact, Quine goes so far as to recommend the adoption of a restrictive axiom in set theory ( $V=L$ ), because it simplifies higher-set theory, noting that simplicity is a crite-

rion of theory acceptance in science. This is despite most set-theorists' rejection of this axiom. It is ironic that Quine, the naturalist, feels comfortable dictating something to mathematicians on philosophical grounds.

**OTHER VERSIONS OF NATURALISM.** Penelope Maddy's (1997) and John Burgess's and Gideon Rosen's (1997) versions of naturalism defend a deferential attitude towards mathematics much like the one Quine shows toward science. They note, first, that mathematics has its own methodology, distinct from so-called scientific method, and that this methodology has proven successful over the centuries. The success of mathematics is measured in mathematical, not scientific terms. Moreover, if mathematicians gave serious pursuit only to those branches known to have applications in natural science, much of the mathematics known in the twenty-first century would not exist, nor would the *science*. The history of science is full of cases where branches of pure mathematics eventually found application in science (see Steiner 1997). That is to say, the overall goals of the scientific enterprise have been well-served by mathematicians pursuing their own disciplines with their own methodology, ignoring science if necessary. Thus, one does not need a direct inferential link between a piece of mathematics and sensory experience before accepting the mathematics as a legitimate part of the web.

On general naturalistic grounds, Burgess and Rosen adopt a realism in ontology for mathematics. For them, the convenience of the face value reading of mathematical propositions counts in its favor. Someone who proposes a nominalistic reconstruction must defend their account on accepted scientific, or mathematical grounds. That is, they must show that the ontology-free versions of mathematics are better mathematics and/or better science. Foregoing philosophical puzzles concerning epistemology do not count. Maddy is more circumspect, arguing that naturalism does not demand a realist interpretation of mathematics.

The varieties of naturalism treated here might be dubbed *methodological* because they focus on the methods of science, adopting those to traditional philosophical questions. Nominalism, as construed here, is an expression of another, ontological variety of naturalism. The thesis is that the only things that exist are the material objects of science, and the only properties people need to consider are the material properties of those objects. Alternately, the only objects in which people are licensed to believe are those with which they causally interact. Mark Colyvan (2001, ch. 3) calls this the *eleatic*

principle. Another issue that separates naturalists—or at least philosophies that go by that name—is whether all legitimate knowledge is empirical. In the spirit of radical empiricism, in the manner of John Stuart Mill, Quine has launched a sustained attack on a priori knowledge. Not every contemporary naturalist follows suit. Bernard Linky and Edward Zalta (1995) argue that the proper interpretation of science requires a more traditional Platonism, according to which mathematical propositions are synthetic a priori. Clearly, an article such as this can do no more than scratch the surface of these rich and wonderful topics.

**See also** Mathematics, Foundations of; Nominalism, Modern.

### Bibliography

- Balaguer, M. *Platonism and Anti-Platonism in Mathematics*. Oxford: Oxford University Press, 1998. Account of realism in ontology and its rivals.
- Benacerraf, P. “What Numbers Could Not Be.” *Philosophical Review* 74 (1965): 47–73. Reprinted in Benacerraf and Putnam, *Philosophy of Mathematics*, 272–294. One of the most widely cited works in the field; argues that numbers are not objects, and introduces an eliminative structuralism.
- Benacerraf, P. “Mathematical Truth.” *Journal of Philosophy* 70 (1973): 661–679. Reprinted in Benacerraf and Putnam, *Philosophy of Mathematics*, 403–420. Another widely cited work; argues that realism in ontology has formidable epistemological problems.
- Benacerraf, P., and H. Putnam, eds. *Philosophy of Mathematics*. 2nd ed. Cambridge, U.K.: Cambridge University Press, 1983. A far-reaching collection containing many of the central articles.
- Brouwer, L. E. J. “Consciousness, Philosophy and Mathematics.” 1948. In *Philosophy of Mathematics: Selected Readings*. 2nd ed., edited by P. Benacerraf and H. Putnam, 90–96. New York: Cambridge University Press, 1983.
- Brouwer, L. E. J. *Intuitionisme en Formalisme*. Groningen, Germany: Noordhof, 1912. Translated as “Intuitionism and Formalism.” In *Philosophy of Mathematics: Selected Readings*. 2nd ed., edited by P. Benacerraf and H. Putnam, 77–79. New York: Cambridge University Press, 1983.
- Burgess, J. “Why I Am Not a Nominalist.” *Notre Dame Journal of Formal Logic* 24 (1983): 93–105. Early critique of nominalism.
- Burgess, J., and G. Rosen. *A Subject With No Object: Strategies for Nominalistic Interpretation of Mathematics*. Oxford: Oxford University Press, 1997. Extensive articulation and criticism of nominalism.
- Chihara, C. *Constructibility and Mathematical Existence*. Oxford: Oxford University Press, 1990. Defense of a modal view of mathematics, and sharp criticisms of several competing views.
- Chihara, C. *A Structural Account of Mathematics*. Oxford: Oxford University Press, 2004. Sequel to Chihara, *Constructibility and Mathematical Existence*, with particular focus on the application of mathematics.
- Colyvan, M. *The Indispensability of Mathematics*. Oxford: Oxford University Press, 2001. Elaboration and defense of the indispensability argument for ontological realism.
- Dummett, M. *Elements of Intuitionism*. Oxford: Oxford University Press, 1977.
- Dummett, M. “The Philosophical Basis of Intuitionistic Logic.” In *Truth and Other Enigmas*. Cambridge, MA: Harvard University Press, 1978. Reprinted in Benacerraf and Putnam, *Philosophy of Mathematics*, 97–129; in Hart, *The Philosophy of Mathematics*, 63–94. Influential defense of intuitionism.
- Dummett, M. *Elements of Intuitionism*. Oxford: Oxford University Press, 1977. Detailed introduction and defense of intuitionistic mathematics.
- Field, H. *Science Without Numbers*. Princeton, NJ: Princeton University Press, 1980. A widely cited defense of fictionalism, by attempting to refute the indispensability argument.
- Field, H. *Realism, Mathematics and Modality*. Oxford: Blackwell, 1989. Reprints of Field’s articles on fictionalism.
- Frege, G. *Die Grundlagen der Arithmetik*. Breslau, Germany: Koebner, 1884. Translated by J. Austin as *The Foundations of Arithmetic*. 2nd ed. New York: Harper, 1960. Classic articulation and defense of logicism.
- Frege, G. *Grundgesetze der Arithmetik*. Vol. 1. Hildesheim, Germany: Olms, 1893. More technical development of Frege’s logicism.
- Gödel, K. “Russell’s Mathematical Logic.” 1944. In *Philosophy of Mathematics*, edited by P. Benacerraf and H. Putnam, 447–469. New York: Cambridge University Press, 1983. Much-cited defense of realism in ontology and realism in truth value.
- Gödel, K. “What is Cantor’s Continuum Problem.” 1964. In *Philosophy of Mathematics*, edited by P. Benacerraf and H. Putnam, 470–485. New York: Cambridge University Press, 1983. Much-cited defense of realism in ontology and realism in truth value.
- Hale, Bob. *Abstract Objects*. Oxford: Basil Blackwell, 1987. Detailed development of neo-logicism, to support Wright (1983).
- Hart, W. D., ed. *The Philosophy of Mathematics*. Oxford: Oxford University Press, 1996. Collection of articles published elsewhere.
- Hellman, G. *Mathematics Without Numbers*. Oxford: Oxford University Press, 1989. Articulation and defense of modal structuralism.
- Heyting, A. *Intuitionism: An Introduction*. Amsterdam: North Holland, 1956. Readable account of intuitionism.
- Hodes, H. “Logicism and the Ontological Commitments of Arithmetic.” *Journal of Philosophy* 81 (1984): 123–149. Another roughly Fregean logicism.
- Linky, B., and E. Zalta. “Naturalized Platonism versus Platonized Naturalism.” *Journal of Philosophy* 92 (1995): 525–555.
- Maddy, P. *Naturalism in Mathematics*. Oxford: Oxford University Press, 1997. Lucid account of naturalism concerning mathematics, and its relation to traditional philosophical issues.
- Maddy, P. *Realism in Mathematics*. Oxford: Oxford University Press, 1990. Articulation and defense of realism about sets.

- Maddy, P. "Three Forms of Naturalism." In Shapiro, *Oxford Handbook of Philosophy of Mathematics and Logic*, 437–459. Contrast with Quinean naturalism.
- Parsons, C. "The Structuralist View of Mathematical Objects." *Synthese* 84 (1990): 303–346. Reprinted in *The Philosophy of Mathematics*, edited by W. D. Hart, 272–309. Oxford: Oxford University Press.
- Putnam, H. *Philosophy of Logic*. New York: Harper Torchbooks, 1971. Source for the indispensability argument for ontological realism.
- Quine, W. V. O. *Ontological Relativity and Other Essays*. New York: Columbia University Press, 1969.
- Quine, W. V. O. *Theories and Things*. Cambridge, MA: Harvard University Press, 1981.
- Resnik, M. *Mathematics as a Science of Patterns*. Oxford: Oxford University Press, 1997. Full articulation of a realist-style structuralism.
- Schirn, M., ed. *Philosophy of Mathematics Today*. Oxford: Oxford University Press, 1998. Proceedings of a conference in the philosophy of mathematics, held in Munich in 1993; coverage of most of the topical issues.
- Shapiro, S., ed. *Oxford Handbook of Philosophy of Mathematics and Logic*. Oxford: Oxford University Press, 2005. Up-to-date articles covering the discipline.
- Shapiro, S. "Philosophy of Mathematics." In *Philosophy of Science Today*, edited by Peter Clark and Katherine Hawley, 181–200. Oxford: Oxford University Press, 2003.
- Shapiro, S. *Philosophy of Mathematics: Structure and Ontology*. Oxford: Oxford University Press, 1997. Extensive articulation and defense of structuralism.
- Shapiro, S. *Thinking about Mathematics: The Philosophy of Mathematics*. Oxford: Oxford University Press, 2000. Popularization and textbook in the philosophy of mathematics.
- Steiner, M. *The Applicability of Mathematics as a Philosophical Problem*. Cambridge, MA: Harvard University Press, 1997.
- Tennant, N. *Anti-Realism and Logic*. Oxford: Oxford University Press, 1987. Articulation of antirealism in truth value, realism in ontology; defends intuitionistic relevance logic against classical logic.
- Tennant, N. *The Taming of the True*. Oxford: Oxford University Press, 1997. Detailed defense of global semantic antirealism.
- Weir, A. "Naturalism reconsidered." In *Oxford Handbook of Philosophy of Mathematics and Logic*, edited by S. Shapiro, 460–482. Oxford: Oxford University Press.
- Whitehead, A. N., and B. Russell. *Principia Mathematica*. Vol. 1. Cambridge, U.K.: Cambridge University Press, 1910.
- Wright, C. *Frege's Conception of Numbers as Objects*. Aberdeen, Scotland: Aberdeen University Press, 1983. Revival of Fregean logicism.

Stewart Shapiro (2005)

## REALITY

See *Appearance and Reality; Being*

## REASON

In English the word *reason* has long had, and still has, a large number and a wide variety of senses and uses, related to one another in ways that are often complicated and often not clear. However, there is one particular sense of the word in which it, with its synonyms or analogues in other languages, has figured prominently in philosophical controversy. This is the sense, sometimes distinguished typographically by an initial capital, in which the term is taken to designate a mental faculty or capacity—in which reason might, for example, be regarded as coordinate with, but distinguishable from, sensation, emotion, or will.

### QUESTIONS TO BE EXAMINED

The question that has been chiefly debated by philosophers might be expressed succinctly, but far from clearly, as "What can reason do?" However, there has also been discussion of the question whether the faculty of reason is peculiar to humanity (and presumably to "higher" beings, if there are any), or whether its possession and exercise in some degree can also be ascribed to "lower" animals. It should perhaps be added that in recent years there has been much debate as to whether machines can, or in principle ever could, properly be said to think; for if an affirmative answer were to be given to this question, then there is a quite common sense of *reason* in which it would follow that that faculty could be exercised by a machine. Only the first of these questions is dealt with here.

The short but unclear question "What can reason do?" is peculiarly liable to give rise to theoretical dissension. The question may, however, be transformed with advantage into a question not directly about the "faculty" of reason itself but about those beings to whom this faculty is attributed. What, we may ask, are human beings in a position to do, in virtue of their possession of the faculty of reason? What, by means of reasoning, are we in a position to achieve? In this form it becomes very clear that the question raises at least two highly disputable issues. First, it is far from immediately clear what reasoning is—on what occasions, in what activities or processes, reason is exercised. And second, if we determine—probably with some degree of arbitrariness—what reasoning is, it may very well remain highly disputable whether this or that can or cannot be achieved by reasoning. One should, indeed, distinguish further at this point between two radically different kinds of dispute that may arise; if it were held that, for instance, knowledge of God cannot be

attained by reasoning, there would plainly be an important further distinction between holding this to be true in fact and true in principle. It might be maintained that the reasoning necessary for knowledge of God is, as a matter of fact, too difficult for frail and mortal human beings to manage; or it might be maintained, quite differently, that the kind of conclusion capable of being established by reasoning excludes in principle that kind, if there is any such, to which knowledge of God must belong. This sort of distinction can be seen as differentiating the positivism preached by Comte in the nineteenth century from the logical positivism of recent philosophy.

### MANY SENSES OF REASON

What, then, is reason? Alternatively, what is reasoning? It seems scarcely possible to maintain that these questions can be given definite answers. The definitions, implicit or explicit, of the relevant terms that have been employed by philosophers and other writers vary widely and significantly; and while some may be judged preferable to others, or may adhere more closely than others to senses which the terms may bear in ordinary discourse, there seems to be no basis secure enough to support a pronouncement that a particular meaning, and hence a particular answer to the question, is exclusively correct. In any case, what is important to the understanding of philosophical writing on this topic is not that one should know what *reason* means but, rather, that one should discern, so far as possible, what meaning is attached to *reason* by an author.

### CONTRASTS WITH OTHER TERMS

Here it seems particularly important and helpful to consider with what reason is contrasted, or from what it is distinguished. There is, for example, a large body of literature in which reason stands essentially in contrast with faith. In this context, what we can achieve by reason is taken to embrace the entire field of knowledge and inquiry in which, with varying degrees of skill and success, we produce or seek reasons for our views, proofs or evidence for our conclusions, and grounds for our opinions. This whole field is set in contrast with another, in which supposedly we may—or should or must—accept certain propositions or doctrines without any grounds but rather on authority or perhaps on unreasoned conviction.

There is another large body of literature in which reason stands in contrast with experience. In this context, what we can achieve by reason is much more narrowly circumscribed; here a distinction is being made between,

roughly, what we can discover or establish by merely sitting and thinking, and what we can discover or establish only by the use of our senses, by observation or by experiment. It will be observed that there are, corresponding to these wider and narrower senses of *reason*, also wider and narrower senses of the term *rationalist*; a rationalist in the one sense is concerned with denying or belittling the claims or the role of faith, and in the other with denying or belittling the role, in the acquisition of knowledge, of experience. There is no particular reason why one who is a rationalist in either one of these senses should be expected to be a rationalist in the other sense also; the two positions are quite independent of one another.

### THE OBJECTS OF REASON

There is, then, no universally agreed or uniquely correct sense of *reason*. This is obvious enough, perhaps; but it is not unimportant. Clearly, even though philosophers may use this term in diverse senses without being wrong, the fact that they do so must, if unobserved by them or their readers, generate confusion and argument at cross purposes. Further, as was noted above, even if we avoid confusion at this point, many problems as to the “scope” or the “powers” of reason remain. They are, in fact, some of the major and central problems of philosophy.

Suppose that, following Brand Blanshard in his *Reason and Analysis*, we define *reason* as “the faculty and function of grasping necessary connections.” We may feel that this is not a very good definition, since it seems excessively restrictive. For example, a judge arguing his way to a decision, or a meteorologist setting forth his grounds for a weather forecast, would in this sense not be exercising the faculty of reason; the argument in each case is nondemonstrative—that is, it does not set out or rely on strictly necessary connections. However, waiving that point, the definition is at least a clear one. But notwithstanding its possession of the important virtue of clarity, the question of what reason can do is not thereby settled.

In order to settle this question, we must decide what necessary connections there are and in what cases or what fields there are necessary connections to be grasped; and the determination of this question raises, or might very well raise, almost every problem of philosophy. Are we to hold, with Plato, that no necessary connections are to be discerned in the everyday world, but only in an intelligible world of Forms? Or are we to hold, with David Hume and many others, that strictly necessary connections are to be found only in the formal, abstract relations between our concepts or ideas? Was Immanuel Kant right in supposing that the moral law can be demonstrated a priori,

and is therefore necessary? Or, on the contrary, was Hume correct in holding that in the field of moral judgment “reason is the slave of the passions”? Are causal relationships cases of necessary connection? Are they perhaps, as John Locke seems to have held, really cases of necessary connection that in practice, however, we are inveterately unable to grasp as such? And so on.

### BASIC QUESTIONS

The point that emerges here is simply this: Whatever particular definition of the faculty of reason we may, implicitly or explicitly, adopt, it seems unavoidable that it will be attempted thereby to distinguish this faculty from others as being that by the exercise of which we can perceive, or arrive at, truths of some particular kind or kinds; and this kind of truth, or these kinds of truths, will in turn be distinguished from other kinds on logical or epistemological grounds. If so, then the question of what we can actually achieve or come to know by reason unavoidably becomes the question of what propositions are of that kind or those kinds; and this is precisely the question about which, in any field, philosophical controversy may, and characteristically does, arise. Thus the apparently simple question “What can reason do?” is not a neutral question on which otherwise dissentient philosophers may expect to be in agreement. On the contrary, it is very likely that their disagreement consists precisely in their diverse answers to this question. It may further be felt, with justice, that if this innocent-looking question unavoidably raises major philosophical issues concerning the logical and epistemological analysis and classification of propositions, it would probably be advantageous to raise those questions directly and overtly rather than as an only half-acknowledged corollary of a discussion that is ostensibly concerned with a faculty of the mind. There are few modern philosophers who would naturally cast their discussions in this latter idiom.

One final risk of confusion is worth pointing out. It is probably true that in recent philosophy there has been a persistent tendency to narrow the field in which necessities, strictly speaking, are admitted to be found; and also, perhaps more significantly, a persistent tendency to take the awesomeness out of necessity by attempts, more or less successful in various fields, to exhibit necessity as fundamentally derived from the unpuzzling, and perhaps unimposing, phenomenon of tautology. In this sense, then, it can be said that there has been some tendency both to narrow the scope conceded to reason and perhaps also to make reason itself seem less mysterious and grand. In some, this tendency has occasioned considerable

distress: As Bertrand Russell has expressed it, “My intellectual journeys have been, in some respects, disappointing.... I thought of mathematics with reverence, and suffered when [Ludwig] Wittgenstein led me to regard it as nothing but tautologies” (*The Philosophy of Bertrand Russell*, edited by P. A. Schilpp, Evanston, IL, 1946, p. 19).

### EXAMINATION OF REASON’S POWERS

There are several instances in which Russell’s sense of distress has been expressed in curiously bellicose terms. Books have been written in defense of reason, and exponents of the contemporary trend have been castigated as reason’s enemies. But this latter charge, even if there is some sense in which it might be well founded, is peculiarly liable to mislead, and very commonly has misled, those who urge it. One thinks, naturally and rightly, of an enemy of reason as one who is opposed or hostile to the exercise of reason. Such a person might be, for instance, a religious bigot, fearful that reason might shake the obscure foundations of his bigotry; he might be a political or racial fanatic, hostile to the careful weighing of arguments and evidence because he is half conscious that his program or doctrine lacks reasonable grounds; or he might, less malignantly, hold some doctrine about the merits of unreflecting spontaneity, disliking the slow pace, the qualifications and hedging, of rational thought. It is obvious, however, that scarcely any philosopher is, or ever has been, an enemy of reason in this sense.

Nor, to mention a group not uncommonly arraigned on the same charge, is the psychoanalyst. It is a tenet of psychoanalytic theory that reason, the dispassionate consideration of arguments and evidence, is a less conspicuous and influential determinant of the beliefs and the conduct of men than has often been supposed, or than most people might like to admit; but the psychoanalyst does not, as would an enemy of reason, rejoice in this circumstance or seek to aggravate it. Quite the contrary: Recognizing the state of the case as being what, in the light of his evidence, he takes it to be, he deploys his art in the attempt to enable people to become more rational than they would otherwise be. He may be mistaken in his theory and unsuccessful in his practice, but in any case neither in theory nor in practice does he display the least enmity toward reason.

Somewhat similarly, the philosopher who produces an argument against high traditional claims for, or traditional characterizations of, reason is, in so doing, exercising reason to the best of his ability; nor does it occur to him to question the desirability of doing so. Thus, to dissent from rationalism as a philosophical doctrine is cer-

tainly not to disparage reason; the man who values, and shows that he values, reason is not he who merely pitches reason's claims exceptionally high but, rather, he who attempts, by painstaking reasoning, to determine how high those claims may justifiably be pitched. Philosophers, whose work consists mostly in sitting and thinking, have often enough and naturally enough been prone to estimate very highly the range and significance of the results that can thereby be achieved. However, this propensity is scarcely an indication of devotion to reason; rather, it is an indication, if of anything, of pardonable self-importance.

**See also** Blanshard, Brand; Comte, Auguste; Faith; Hume, David; Locke, John; Logical Positivism; Plato; Positivism; Practical Reason; Rationalism; Russell, Bertrand Arthur William; Thinking; Wittgenstein, Ludwig Josef Johann.

### **Bibliography**

- Bermudez, Jose Luis, and Alan Millar, eds. *Reason and Nature: Essays in the Theory of Rationality*. Oxford: Clarendon Press, 2002.
- Blanshard, Brand. *Reason and Analysis*. London: Allen and Unwin, 1962.
- BonJour, Laurence. *In Defense of Pure Reason*. New York: Cambridge University Press, 1997.
- Brewer, Bill. *Perception and Reason*. Oxford: Clarendon Press, 2002.
- Chisholm, Roderick. *Theory of Knowledge*. Englewood Cliffs, NJ: Prentice-Hall, 1966; 2nd ed., 1977; 3rd ed., 1989.
- Ewing, A. C. *Reason and Intuition*. London: H. Milford, 1942.
- Foley, Richard. *The Theory of Epistemic Rationality*. Cambridge, MA: Harvard University Press, 1987.
- Freud, Sigmund. *The Future of an Illusion*. Translated by W. D. Robson-Scott. London: Hogarth Press, 1928.
- Friedman, Michael. "Philosophy as Dynamic Reason: The Idea of a Scientific Philosophy." In *What Philosophy Is: Contemporary Philosophy in Action*, edited by Havi Carel. London: Continuum, 2004.
- Howson, Colin, and Peter Urbach. *Scientific Reasoning: The Bayesian Approach*. 2nd ed. Chicago: Open Court, 1993.
- Moser, Paul K., ed. *Rationality in Action*. Cambridge, U.K.: Cambridge University Press, 1990.
- Murphy, Arthur E. *The Uses of Reason*. New York: Macmillan, 1943.
- Nagel, Ernest. *Sovereign Reason*. Glencoe, IL: Free Press, 1954.
- Papineau, David. *The Roots of Reason: Philosophical Essays on Rationality, Evolution, and Probability*. Oxford: Clarendon Press, 2003.
- Peacocke, Christopher. *The Realm of Reason*. Oxford: Clarendon Press, 2004.
- Russell, Bertrand. *Skeptical Essays*. New York: Norton, 1928.
- Santayana, George. *The Life of Reason*, rev. ed. New York: Scribners, 1954.

Stich, Stephen. *The Fragmentation of Reason*. Cambridge, MA: MIT Press, 1990.

Walsh, W. H. *Reason and Experience*. Oxford: Clarendon Press, 1947.

Whitehead, A. N. *The Function of Reason*. Princeton, NJ: Princeton University Press, 1929.

**G. J. Warnock (1967)**

*Bibliography updated by Benjamin Fiedor (2005)*

## REASONING

See *Thinking*

## REBIRTH

See *Reincarnation*

## RECURSION THEORY

See *Computability Theory*

## REDUCIBILITY, AXIOM OF

See *Russell, Bertrand (section on logic and mathematics)*

## REDUCTION

A cursory glance at the history of science reveals a continuous succession of scientific theories of various areas or domains. For example, since ancient times theories of the cosmos have been proposed to account for the observed behavior of the heavenly bodies. The geocentric Ptolemaic theory was, for instance, succeeded by the heliocentric theory of Copernicus. Another example concerns the nature of light. Corpuscular theories were succeeded by wave theories of light. Wave theories, in turn, have been followed by the quantum theories of electromagnetic radiation.

This entry concerns the nature of certain relations that may obtain between different pairs of theories in such sequences. A radical or extreme view of those relations is that of Thomas Kuhn. Kuhn (1970) famously argues that across scientific revolutions there is a radical disconnect between theories. One can find a similar argument in Paul K. Feyerabend (1962). On such a view, no

rational relations can obtain between a theory and its predecessor. However, it is fair to say that most philosophers have held, contrary to this extreme position, that there are, indeed, interesting and contentful relations between various pairs of scientific theories. One such relationship is that of reduction. It is often claimed that successor theories reduce those that they succeed. Such a relation may involve the idea that the successor or reducing theory explains or otherwise absorbs the successful features of the reduced theory. However, getting clear about exactly how the notion of reduction should be understood has been and continues to be a difficult philosophical problem.

This entry begins with a discussion of what may be called the received view of theory reduction and examines how that view has evolved as the result of various criticisms. Work on intertheoretic relations from 1997 through 2005 is then considered.

### NAGELIAN REDUCTION

The *locus classicus* for contemporary discussion is Ernest Nagel's presentation of a model for theory reduction in *The Structure of Science* (1961). Nagel takes reduction to be an explanatory relation between theories where explanation is understood to involve deductive logical relations between statements characterizing the explanans and the statement characterizing the explanandum in accordance with the Hempelian (Hempel 1965) deductive-nomological model. Nagel holds that "[r]eduction ... is the explanation of a theory or a set of experimental laws established in one area of inquiry, by a theory usually though not invariably formulated for some other domain" (1961, p. 338). The idea here is that a theory *T* reduces a theory *T'* just in case one can derive (and thereby explain) the laws of *T'* from the laws of *T*.

Nagel realizes that for some intuitive cases of theory reduction such derivations would not be immediately possible. If the vocabulary of the reduced (succeeded) theory contains terms referring to entities or properties that are not mentioned in the vocabulary of the reducing (successor) theory, then it will be impossible to derive the laws of the reduced theory containing those terms from the laws of the reducing theory. Reductions involving theories with distinct vocabularies are called heterogeneous by Nagel. By contrast, homogeneous reductions are taken by him to be rather straightforward and unproblematic.

This view of homogeneous reductions is somewhat naive. Lawrence Sklar (1967) points out that homogeneous reductions, in fact, are rare. Instead, what one has typically is the derivation of an approximation to the

reduced theory and not of the reduced theory itself. An example discussed by both Nagel and Sklar concerns the homogeneous reduction of the Galilean theory of free fall to Newtonian mechanics and gravitational theory. Sklar notes that there really is no strict derivation of the Galilean theory, although no terms appear in the Galilean theory that do not also appear in Newton's theory.

The example of a heterogeneous reduction Nagel discusses is the apparent reduction of thermodynamics to statistical mechanics. This example has become paradigmatic of intertheoretic reduction in the general philosophical literature. (In actual fact, the reduction of thermodynamics to statistical mechanics is much more complex than Nagel's discussion allows. Sklar [1993] provides a detailed discussion of various difficulties involved in the reduction of thermodynamics to statistical mechanics.) Thermodynamics contains terms referring to properties such as temperature and entropy. Such terms are completely lacking in the vocabulary of statistical mechanics. To effect the (supposed) derivational reduction, one must connect these thermodynamic terms with terms occurring in the vocabulary of statistical mechanics.

Nagel introduces two necessary formal conditions for such heterogeneous reductions:

- *Connectability*: "Assumptions of some kind must be introduced which postulate suitable relations between whatever is signified by 'A' [a term appearing in the reduced but not the reducing theory, such as 'temperature'] and traits represented by theoretical terms already present in the primary [reducing] science."
- *Derivability*: "With the help of these additional assumptions, all the laws of the secondary [reduced] science, including those containing the term 'A,' must be logically derivable from the theoretical premises and their associated coordinating definitions in the primary [reducing] discipline." (1961, pp. 353–354)

The connectability requirement is vague as it stands. What is the exact nature of the required "suitable relations"? In the literature such relations of connectability are typically called bridge laws or bridging hypotheses and their status is a matter of debate. Nagel allows that such bridge laws need not have the form of universally quantified biconditionals for theory reduction to be possible. They might, he holds, have the form of one-way conditionals. It is this possibility that renders the requirement of derivability not superfluous (Nagel 1961, p. 355

note). With the aid of bridge laws, Nagel thinks that the reducing theory would be able to fully explain the laws of the reduced theory.

However, even having universal biconditionals as bridge laws may not itself be sufficient for reduction. Many examples exist where correlatory laws may be established—where the biconditionals are true and apparently lawlike—yet, where nothing resembling reduction can take place. Sklar (1967) offers the example of the Wiedemann-Franz law expressing a correlation between the thermal conductive properties of a material and its electrical conductivity properties. Such a law does not allow one to reduce the theory of thermal conductive properties of the material to a theory of its electrical conductive properties. Something more than mere correlation is required.

That something more is usually taken to be some kind of empirically established identity claim. For example, the reduction of physical optics to the theory of electromagnetic radiation is accomplished by noting the identity of one class of entities—light waves—with (part of) another class—electromagnetic radiation. As Sklar notes, “Light waves are not correlated with electromagnetic waves, for they are electromagnetic waves. There are not two classes of entities, but only one” (1967, p.120). Another classic example is the reduction of Mendelian genetics to molecular genetics via the identification of genes with DNA molecules.

The idea that the bridge laws must express necessary identifications between entities or classes of entities has much to recommend it. However, in many cases of apparent intertheoretic reduction such identity relations are not available. In the paradigmatic case of the reduction of thermodynamics to statistical mechanics, one sees that terms such as *temperature* and *entropy*, occurring in thermodynamics but not in statistical mechanics, refer to properties possessed by thermodynamic systems. Still, it is not at all clear what properties of statistical systems can be identified with the thermodynamic properties. For example, the standard claim that temperature is just (identical to) mean molecular kinetic energy is deeply problematic. Again, see Sklar (1993) for a detailed discussion of some of these problems.

One way of emphasizing the difficulty here is in terms of questions about the meaning of terms appearing in the distinct theories. In orthodox thermodynamics, for example, the term *entropy* gets its meaning (on one view of how theoretical terms acquire meaning) at least in part by the role the term plays in the theory. (A classic presentation of orthodox thermodynamics explicitly exhibiting

the roles of the terms is by A. B. Pippard [1957].) One sees that such terms refer to unvarying and nonstatistical properties of systems. Nevertheless, in the apparent reduction of thermodynamics to statistical mechanics the concept of entropy changes to one that explicitly allows for statistical variation and fluctuation. In what sense can one identify here? Feyerabend (1962), for one, takes this to be evidence that reduction (understood as Nagelian derivation with bridge laws) must fail.

**NEO-NAGELIAN REDUCTION.** In contrast to Feyerabend’s (1962) pessimistic conclusion many philosophers hold that some sort of reductive relation still obtains even in the face of problems of heterogeneity. In fact, it is often noted that in the process of reducing one theory to another, the reduced theory gets emended. One sees textbooks with titles referring to statistical thermodynamics, indicating that the orthodox thermodynamic conceptions of entropy and temperature have been changed to allow for (observable and observed) fluctuations in those quantities. The explicit recognition that the reduced theory is often changed as a result of reduction or attempted reduction takes one beyond the Nagelian conception of reduction as a relatively straightforward explanatory derivation.

Kenneth Schaffner’s (1967, 1976) model of reduction deserves mention here as a sophisticated attempt to incorporate this aspect of theoretical change into a Nagelian-type framework. Schaffner explicitly includes the corrected reduced theory in the model. On this view a theory  $T$  reduces a theory  $T'$  just in case there is a corrected version of the reduced theory,  $T'^*$  such that

- (1) The primitive terms of  $T'^*$  are associated with various terms of  $T$  via bridge laws or reduction functions
- (2)  $T'^*$  is derivable from  $T$  when supplemented by these bridge laws
- (3)  $T'^*$  corrects  $T$  in that it makes more accurate predictions than does  $T'$
- (4)  $T'$  is explained by  $T$  in that  $T'$  and  $T'^*$  are strongly analogous to one another, and  $T$  says why  $T'$  works as well as it does in its domain of validity.

It is clear that work must be done to explicate the intuitive notion of strong analogy playing a role in this model of reduction. See William C. Wimsatt (1976) for some suggestions along these lines.



## OBJECTIONS TO NAGELIAN REDUCTIONS

A number of influential objections have been raised against Nagelian models of reduction. Most of these concern the possibility of providing the appropriate bridge laws. As a result they can be seen as telling also against more sophisticated models such as Schaffner's (1967, 1976). Additionally, it has been objected that even if such bridge laws can be provided, there remains an explanatory question about their status as laws. Consider the second objection first.

**EXPLANATORY QUESTIONS ABOUT BRIDGE LAWS.** In those cases where bridge laws express the identification of classes of entities, to ask why those bridge laws hold is to ask a question that can be trivially answered. The reason the bridge laws hold is because the entities in question are one and the same. "Why should I believe that light waves are electromagnetic radiation?" Answer: "They just are. Period, end of story." By contrast, in cases where bridge laws express some kind of (perhaps, nomologically) necessary coextensivity between properties appearing in two theories, such a question may seem legitimate and answers may be hard to come by. Jaegwon Kim (1998) forcefully argues that this poses a serious problem for Nagelian reduction understood as attempting to effect an explanatory relation among pairs of theories.

Kim discusses the attempted Nagelian reduction of psychology (the science of the mental) to physical theory, say, neurophysiology. One can suppose that one discovers empirically a nomological correlation between being in pain and having one's C-fibers firing. A statement characterizing this correlation is taken to be a bridge law necessary for Nagelian reduction. In this case it seems reasonable to ask: "Can we understand why we experience pain when our C-fibers are firing, and not when our A-fibers are firing? Can we explain why pains, not itches or tickles, correlate with C-fiber firings?" (Kim 1998, p. 95). Kim's point is that if Nagelian reduction is supposed to provide an explanation of the reduced theory in terms of the reducing theory, then surely one must demand an explanation of the bridge laws employed in the explanatory derivation. "*For it is the explanation of these bridge laws, an explanation of why there are just these mind-body correlations, that is at the heart of the demand for an explanation of mentality*" (p. 96, emphasis in the original).

**MULTIPLE REALIZABILITY.** A different argument due to Jerry Fodor (1974) has been used to block attempts at almost every Nagelian reduction of a given (special sci-

ence) theory to more basic (physical) theory. This argument has come to be called the multiple realization argument. It depends on the assumption that properties appearing in the special (to-be-reduced) science may have diverse and "wildly heterogeneous" realizers in the reducing physical theory. As Fodor puts it, "The problem ... has been that there is an open empirical possibility that what corresponds to the natural kind predicates of a reduced science may be a heterogeneous and unsystematic disjunction of predicates in the reducing science" (p. 108). Thus, to continue the psychology example so prevalent in the literature, pain—a property appearing in the science of psychology whose predicate (perhaps) appears in its laws—may be realized by distinct physical or neurophysiological properties in humans, in reptiles, and possibly even in inorganic robots.

This has the consequence no one neurophysiological state can be correlated or identified with the psychological property pain. In humans it may be C-fibers firing; in reptiles it may be D-fibers firing; and in robots it may be the activation of some particular integrated circuit. The heterogeneous nature of the distinct realizers also makes it unlikely that a disjunction of those realizers will be a natural kind term in the reducing theory. Given this, and if laws relate natural kinds to natural kinds, it is unlikely that there can be anything lawlike about the bridge laws. This argument has been applied to many functionally defined properties such as being a thermostat or being a heart—properties that can be realized in many different ways in different systems or organisms.

One response, due to Kim (1992), to the realization argument is to note that while the argument may block a kind of global reduction of the special science to the lower-level physical theory, it may be possible to have (local) species or structure specific reductions. Thus, for instance, one might be able to locally reduce human pain to human neurophysiology, reptilian pain to reptilian neurophysiology, and robot pain to robot neurocircuitry. Kim (1998, chapter 4) develops an alternative functional model of reduction appropriate to this response.

Another approach (Batterman 2000) asks for an account of what makes the multiple realizability possible. Typically, multiple realizability is simply assumed and applied via the multiple realization argument to block Nagelian reductions. For instance, Fodor (1974) cites it simply as an open possibility. However, it seems reasonable to ask whether one can explain, from the point of view of the supposed reducing theory, the possibility of multiple realizability. If so, this may lead to a kind of explanation without reduction. Cases where such expla-

nations are indeed possible can be found in the physics literature where attempts are made to explain surprising universal features of various systems. *Universality* means identical or similar behavior in physically distinct systems and is, therefore, a term essentially synonymous with *multiple realizability*. For details about how such an approach to multiple realizability will go, see Robert W. Batterman (2000, 2002).

## REDUCTION IN THE OTHER DIRECTION

There is an interesting terminological ambiguity that infects the term *reduction* as it is typically used in the philosophical literature and as it is used in the physics literature. Philosophers will discuss the reduction of thermodynamics to statistical mechanics, the reduction of classical mechanics to quantum mechanics, and the reduction of the ray theory of light to the wave theory. The succeeded theories are all reduced to their successors. Physicists, when they talk about theory reduction at all, tend to put things the other way around. They will say that statistical mechanics reduces to thermodynamics in the limit as the number of degrees of freedom goes to infinity. They will say that quantum mechanics reduces to classical mechanics in some kind of correspondence limit. Furthermore, they will say that the wave theory reduces to the ray theory in the limit as the wavelength of light approaches zero. That there are such different senses of intertheoretic reduction was first noted by Thomas Nickles in the paper “Two Concepts of Intertheoretic Reduction” (1973).

Interestingly, the physicists’ sense of reduction appeals to limiting relations between the pair of theories and is not concerned with derivation in logical/syntactic sense that has primarily concerned philosophers following in Nagel’s footsteps. In other words, there is no explicit concern, say, with the derivation of the laws of thermodynamics from the laws of statistical mechanics. Instead, the interest is in the potential emergence of those laws as some sort of mathematical limit is asymptotically approached. Thus, while the philosophical tradition focuses on the schema according to which theory  $T'$  reduces to theory  $T$  just in case one can derive (and thereby explain) the laws of  $T'$  from the laws of  $T$ , this other sense of reduction focuses on a schema of the following form (1):

$$\lim_{\epsilon \rightarrow 0} T = T'$$

in which theory  $T$  reduces to  $T'$  in the regime where a parameter  $\epsilon$  appearing in theory  $T$  takes on a limiting value. For instance, quantum mechanics contains a constant (Planck’s constant) that plays no role in classical mechanics. As a result, one may be motivated to study the limit of quantum mechanics in which Planck’s constant approaches zero. This is a kind of correspondence limit.

The two schemas are related to one another at least in the following way. Should the equality in (1) hold for two theories  $T$  and  $T'$ , then it is reasonable to expect that the laws of  $T'$  are derivable from those of  $T$ . That is to say, it is likely that one will be able to find the appropriate connections that will allow something like a Schaffner-style neo-Nagelian reduction. On the contrary, if the equality in (1) fails to obtain for the pair of theories, then such neo-Nagelian reduction will not be possible. It will be impossible to form the relevant corrected reduced theory  $T'^*$ .

One case for which the schema (1) does obtain is in the relationship between (certain aspects of) classical Newtonian mechanics (NM) and the special theory of relativity (SR). In the limit in which velocities are slow compared with the speed of light ( $(v/c) \rightarrow 0$ ), SR reduces to NM. The limit exists and the formulas of SR smoothly (that is uniformly) approach those of NM.

However, far more often than not pairs of theories will be related to one another by so-called singular limits. Singular limits arise when the behavior as the limit is approached (no matter how small  $\epsilon$  becomes) is qualitatively different from the behavior at the limit (when  $\epsilon = 0$ ). In such cases the equality in schema (1) fails to obtain: There will be no smooth approach of the formulas of theory  $T$  to those of theory  $T'$ .

In fact, it is fair to say that for most theory pairs of interest, schema (1) will fail. Important examples include those mentioned earlier: quantum mechanics and classical mechanics; wave theory and ray theory; and statistical mechanics and thermodynamics. Certain formulas in each of these theory pairs are related by singular limits, and it is best, perhaps, to give up on speaking of reductive relations between the theories or at least between those features characterized by the singularly related formulas (see Berry 1994, 2002; Batterman 1995, 2002).

It is important to stress that if this is correct, and so physicists’ reductions are genuinely few and far between, this does not mean that there is no reason to study the singular limiting relationships between theories. In fact, the opposite is true. There is much of interest to study in the borderland between the theories. Michael V. Berry

notes the importance of the failure of reduction due to singular limiting relations between the theories, “[M]any difficulties associated with reduction arise because they involve singular limits. These singularities have both negative and positive aspects: They obstruct smooth reduction of more general theories to less general ones, but they also point to a great richness of borderland physics between theories” (2001, p. 42).

The “great richness” of this borderland is fertile ground for studying certain aspects of emergence, a philosophical topic related to the failure of reduction. Emergent phenomena are typically taken to be novel in certain respects where this novelty is often understood as resulting from the failure of the more basic theory to explain or otherwise account for the phenomena. There has been considerable interest in the controversial issues surrounding the nature and existence of emergent phenomena. Two related approaches with the same starting point—the singular nature of limiting intertheoretic relations—are examined by Batterman (2002) and Alexander Rueger (2000a, 2000b). However, another related approach can be found in the work of Hans Primas (1998). For a different, more metaphysically motivated attempt to understand emergence, see Paul Humphreys (1997). This is currently an active area of research.

**See also** Copernicus, Nicolas; Fodor, Jerry A.; Galileo Galilei; Hempel, Carl Gustav; Kuhn, Thomas; Laws and Theories; Multiple Realizability; Nagel, Ernest; Newton, Isaac; Philosophy of Science; Properties; Scientific Theories.

### Bibliography

- Batterman, Robert W. *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. New York: Oxford University Press, 2002.
- Batterman, Robert W. “Multiple Realizability and Universality.” *British Journal for the Philosophy of Science* 51 (2000): 115–145.
- Batterman, Robert W. “Theories between Theories: Asymptotic Limiting Intertheoretic Relations.” *Synthese* 103 (1995): 171–201.
- Berry, Michael V. “Asymptotics, Singularities, and the Reduction of Theories.” In *Logic, Methodology, and Philosophy of Science*, edited by Dag Prawitz, Brian Skyrms, and Dag Westerståhl, 597–607. Amsterdam, Netherlands: Elsevier Science B. V., 1994.
- Berry, Michael V. “Chaos and the Semiclassical Limit of Quantum Mechanics (Is the Moon There When Somebody Looks?).” In *Quantum Mechanics: Scientific Perspectives on Divine Action*, edited by Robert John Russell et al., 41–54. Rome, Italy: Vatican Observatory CTNS Publications, 2001.
- Berry, Michael V. “Singular Limits.” *Physics Today* (May 2002): 10–11.
- Feyerabend, Paul K. “Explanation, Reduction, and Empiricism.” In *Minnesota Studies in the Philosophy of Science. Vol. 3, Scientific Explanation, Space, and Time*, edited by Herbert Feigl and Grover Maxwell. Minneapolis: University of Minnesota Press, 1962.
- Fodor, Jerry. “Special Sciences, or the Disunity of Sciences as a Working Hypothesis.” *Synthese* 28 (1974): 97–115.
- Hempel, Carl G. *Studies in the Logic of Explanation: In Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press, New York, 1965.
- Humphreys, Paul. “How Properties Emerge.” *Philosophy of Science* 64 (1) (1997): 1–17.
- Kim, Jaegwon. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press, 1998.
- Kim, Jaegwon. “Multiple Realization and the Metaphysics of Reduction.” *Philosophy and Phenomenological Research* 52 (1) (1992): 1–26.
- Kuhn, Thomas S. *The Structure of Scientific Revolutions*. 2nd ed. Chicago: University of Chicago Press, 1970.
- Nagel, Ernest. *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace, 1961.
- Nickles, Thomas. “Two Concepts of Intertheoretic Reduction.” *Journal of Philosophy* 70 (7) (1973): 181–201.
- Pippard, A. B. *Elements of Classical Thermodynamics*. Cambridge, U.K.: Cambridge University Press, 1957.
- Primas, Hans. “Emergence in Exact Natural Sciences.” *Acta Polytechnica Scandinavica, Ma* 91 (1998): 83–98.
- Rueger, Alexander. “Physical Emergence, Diachronic and Synchronic.” *Synthese* 124 (2000a): 297–322.
- Rueger, Alexander. “Robust Supervenience and Emergence.” *Philosophy of Science* 67 (2000b): 466–489.
- Schaffner, Kenneth. “Approaches to Reduction.” *Philosophy of Science* 34 (2) (1967): 137–147.
- Schaffner, Kenneth. “Reductionism in Biology: Prospects and Problems.” In *PSA 1974*, edited by A. C. Michalos et al., 613–632. Dordrecht, Netherlands: D. Reidel, 1976.
- Sklar, Lawrence. *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*. New York: Cambridge University Press, 1993.
- Sklar, Lawrence. “Types of Inter-theoretic Reduction.” *British Journal for the Philosophy of Science* 18 (1967): 109–124.
- Wimsatt, William C. “Reductive Explanation: A Functional Account.” In *PSA 1974*, edited by A. C. Michalos et al., 671–710. Dordrecht, Netherlands: D. Reidel, 1976.

Robert W. Batterman (2005)

## REDUCTIONISM IN THE PHILOSOPHY OF MIND

See Appendix, Vol. 10

## REFERENCE

“Reference” is usually conceived as the central relation between language or thought and the world. To talk or think about something is to refer to it. Twentieth-century philosophy found such relations particularly problematic. One paradigm of reference is the relation between a proper name and its bearer. On a more theoretical conception all the constituents of an utterance or thought that contribute to determining whether it is true refer to their contributions (as, for example, a predicate refers to a property). In analytic philosophy discussion of reference was dominated until the 1960s by the views of Gottlob Frege and Bertrand Russell and modifications of them (such as those by P. F. Strawson). Criticisms of assumptions common to those views then provoked a revolution in the theory of reference. The alternatives include causal and minimalist theories.

### OBJECTIONS TO DESCRIPTIVISM

One model of reference is that of descriptive fit. The paradigm is a definite description (such as “the tallest tree”) that refers to whatever it accurately describes. Frege and Russell assimilated the reference of ordinary proper names to this case by supposing that speakers associate them with descriptions. Similar accounts were later given of mass terms (such as “blood”), natural-kind terms (“gorilla”), and theoretical terms in science (“inertia”). It was conceded that most terms are associated with vague and context-dependent clusters of descriptions and that reference might be to whatever they least inaccurately described, but such liberalizations did not challenge the underlying idea that descriptive fit determines reference. However, Keith Donnellan, Saul Kripke, and Hilary Putnam proposed counterexamples to that idea. Suppose, for instance, that speakers associate the name “Jonah” with the Bible story. Traditional descriptivism concludes that the sentence “Although Jonah existed, those things happened only to someone else” is untrue. For if one person satisfied the relevant descriptions, “Jonah” would refer to him. But then descriptivism proves too much, for philosophical reflection cannot show that the Bible story is not a mere legend that grew up about a real person; if those things really happened to someone else, of whom no word reached the biblical writer, the name “Jonah” would still refer to the former, not the latter. Similarly, traditional descriptivism permits someone who thinks of gorillas primarily as ferocious monkeys to conclude falsely that the sentence “Gorillas exist, but they are not ferocious monkeys” is untrue.

A second criticism was this. Say that a term *t* rigidly designates an object *x* if and only if *t* designates (refers to) *x* with respect to all possible circumstances (except perhaps for circumstances in which *x* does not exist). Most descriptions designate nonrigidly: “the tallest tree” designates one tree with respect to present circumstances, another with respect to possible circumstances in which the former is outgrown. The descriptions that traditional descriptivists associated with names were nonrigid. However, names designate rigidly: Although we can envisage circumstances in which the Danube would have been called something else instead, we are still using our name “Danube” to hypothesize circumstances involving the very same river. Thus, most descriptions do not behave like names.

The second criticism was met by a modification of descriptivism. The descriptions associated with a name were rigidified by a qualifying phrase such as “in present circumstances.” “The tallest tree in present circumstances” rigidly designates what “the tallest tree” nonrigidly designates. The first criticism is less easily met. Some descriptivists used deferential descriptions such as “the person referred to in the Bible as ‘Jonah.’” A more general strategy is to exploit the success of any rival theory of reference by building that theory into the associated descriptions. However, such moves jeopardize the connection between reference and speakers’ understanding (a connection that descriptivism was intended to secure) as the descriptions that speakers supposedly associate with names become less and less accessible to the speakers themselves.

It is in any case clear that, as Russell recognized, not all reference is purely descriptive. If the sentence “It is hot now” is uttered at different times in exactly similar circumstances, associated with exactly the same descriptions, those descriptions are not what determines that it changes its reference from one time to the other. The reference of a token of “now” is determined by the time of its production and the invariant linguistic meaning of “now,” the rule that any such token refers to the time of its production. Similarly, the presence of an object to the speaker or thinker plays an ineliminably nondescriptive role in the reference of demonstratives such as “this.”

### NONDESCRIPTIVISM

**THE KRIPKE-PUTNAM PICTURE.** Kripke and Putnam proposed an alternative picture. Something *x* is singled out, usually demonstratively (“this river,” “this kind of animal”). A name *n*, proper or common, is conferred on *x* (“Danube,” “gorilla”). The name is passed on from one

speaker to another, the latter intending to preserve the former's reference. Such intentions are self-fulfilling:  $n$  continues to refer to  $x$ . The beliefs that speakers would express in sentences containing  $n$  play no role in making  $n$  refer to  $x$ , so it can turn out that most of them are false. The picture involves two kinds of deference. Synchronically, there is division of linguistic labor: Ordinary speakers defer to experts (as in deciding which animals "gorilla" refers to). Diachronically, later speakers defer to earlier ones in a historical chain. Thus, reference typically depends on both the natural environment of the initial baptism (to fix the demonstrative reference) and the social environment of the later use. An individual speaker's understanding plays only a minor role. The account may be generalized (as to many adjectives and verbs).

The picture needs qualification. Gareth Evans pointed out that a name can change its reference as a result of misidentification, even if each speaker intends to preserve reference. What matters is not just the initial baptism but subsequent interaction between word and object. Such concessions do not constitute a return to descriptivism.

**CAUSAL THEORIES.** The Kripke-Putnam picture is often developed into a causal theory of reference, on which for  $n$  to refer to  $x$  is for a causal chain of a special kind to connect  $n$  to  $x$ . Such a theory goes beyond the original picture in at least two ways. First, although that picture required later uses of  $n$  to depend causally on the initial baptism, it did not require the initial baptism to depend causally on  $x$ . Kripke allowed reference to be fixed descriptively (not just demonstratively), as in "I name the tallest tree 'Albie'"; he merely insisted that the description did not give the meaning of the name. There is no causal connection between the name "Albie" and the tree Albie. Second, Kripke and Putnam did not attempt to define the notions they used in causal terms; the notion of an intention to preserve reference is not obviously causal.

Causal theories are often motivated by a desire to naturalize linguistic and mentalistic phenomena by reducing them to the terms of physical science. Such theories are therefore not restricted to proper names. Causal theorists will postulate that our use of the words "tall" and "tree" is causally sensitive to tallness and trees respectively, hoping thereby to explain the reference of "Albie." One problem for causal theories is that any word is at the end of many intertwined causal chains with different beginnings. It is extremely difficult to specify in causal

terms which causal chains carry reference. For this reason, causal theories of reference remain programmatic.

**DIRECT REFERENCE.** Consonant with the Kripke-Putnam picture, but independent of causal theories of reference, is the theory of direct reference developed by David Kaplan. A term  $t$  directly refers to an object  $x$  in a given context if and only if the use of  $t$  in that context contributes nothing to what is said but  $x$  itself. For Kaplan, proper names, demonstratives, and indexicals such as "now" refer directly. Ruth Barcan Marcus had earlier made the similar suggestion that proper names are mere tags. The reference of a directly referential term may be determined relative to context by its context-independent linguistic meaning, as for "now"; the claim is that what "now" contributes to the proposition expressed by an utterance of "It is hot now" is not its invariant linguistic meaning but the time itself.

Although all direct reference is rigid designation, not all rigid designation is direct reference: "the square of 7" rigidly designates 49, but the reference is not direct, for the structure of the description figures in the proposition expressed by "The square of 7 is 49." On one view all genuine reference is direct, sentences of the form "The  $F$  is  $G$ " being quantified on the pattern of "Every  $F$  is  $G$ " (as Russell held); "the  $F$ " is neither a constituent nor a referring term.

If "Constantinople" and "Istanbul" have the same direct reference, the proposition (C) expressed by "Constantinople is crowded" is the proposition (I) expressed by "Istanbul is crowded," so believing (C) is believing (I), even if one would not express it in those words. Similarly, when a term of a directly referential type fails to refer, sentences in which it is used express no proposition. The view is anti-Fregean. In suitable contexts Frege would attribute different senses but the same reference to "Constantinople" and "Istanbul" and a sense but no reference to an empty name; for him the sense, not the reference, is part of what is said or thought. Russell held that logically proper names are directly referential but concluded that ordinary names are not logically proper. The challenge to defenders of the direct-reference view is to explain away the appearance of sameness of reference without sameness of thought and absence of reference without absence of thought, perhaps by postulating sense-like entities in the act rather than the content of thought. The theory of direct reference concerns content, not the mechanisms of reference.

**MINIMALISM.** Traditional theorizing about reference is ambitious; the possibility of a broad and deep theory such as it seeks has been questioned by Richard Rorty, Robert Brandom, Paul Horwich, and others. The following schema constitutes a minimal account of reference (“*a*” is replaceable by singular terms):

(R) For any *x*, “*a*” refers to *x* if and only if  $x = a$ .

“London” refers to London and nothing else. A minimalist account adds to (R) the claim that (R) exhausts the nature of reference.

Some qualifications are necessary. First, if anything but a singular term replaces “*a*” in (R), the result is ill formed, for only singular terms should flank the identity sign. If expressions of other syntactic categories refer, those categories will require their own schemas. The schema for predicates might be:

(R') For any *x*, “*F*” refers to *x* if and only if  $x = Fness$ .

Second, the notion of a singular term must be explained (can “my sake” replace “*a*”?). Third, (R) does not say which singular terms refer. When “*a*” does not refer, (R) may not express a proposition. Fourth, (R) cannot be generalized by the prefix “In all contexts”: “today” used tomorrow does not refer to today. Rather, (R) should be understood as instantiated by sentences in different contexts (for instance, uttered tomorrow with “today” for “*a*”). Fifth, when one cannot understand the term “*a*,” one cannot understand (R). Thus, one will find many instances of (R) unintelligible.

One’s grasp of the minimal theory is not a grasp of each of many propositions; it is more like one’s grasp of a general pattern of inference. For (R) the pattern is in the sentences that express the propositions, not in the propositions themselves (it is not preserved when a synonym replaces the unquoted occurrence of “*a*”). This generality does not satisfy all philosophers. Many accept the minimal theory but reject minimalism, because they postulate a deeper (for instance, causal) theory of reference that explains (R) and (R'). Although the reductionist demand for strictly necessary and sufficient conditions for reference in more fundamental terms may be overambitious, a good picture of reference might still reveal more than (R) and (R') without meeting that demand.

**See also** Frege, Gottlob; Indexicals; Kaplan, David; Kripke, Saul; Marcus, Ruth Barcan; Philosophy of Language; Proper Names and Descriptions; Putnam, Hilary; Rorty, Richard; Russell, Bertrand Arthur William; Sense; Strawson, Peter Frederick.

## Bibliography

- Almog, J., J. Perry, and H. Wettstein, eds. *Themes from Kaplan*. New York: Oxford University Press, 1989.
- Brandom, R. “Reference Explained Away: Anaphoric Reference and Indirect Description.” *Journal of Philosophy* 81 (1984): 469–492.
- Devitt, M. *Designation*. New York: Columbia University Press, 1981.
- Evans, G. *The Varieties of Reference*. Oxford: Clarendon Press, 1982.
- Fodor, J. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press, 1987.
- French, P., T. Uehling, and H. Wettstein, eds. *Contemporary Perspectives in the Philosophy of Language*. Midwest Studies in Philosophy, Vol. 5. Minneapolis: University of Minnesota Press, 1979.
- Horwich, P. *Truth*. Oxford: Blackwell, 1990.
- Kripke, S. *Naming and Necessity*. Cambridge, MA: Harvard University Press, 1980.
- Lewis, D. “Putnam’s Paradox.” *Australasian Journal of Philosophy* 62 (1984): 221–236.
- Neale, S. *Descriptions*. Cambridge, MA: MIT Press, 1990.
- Putnam, H. *Philosophical Papers*, Vol. 2: *Mind, Language, and Reality*. Cambridge, U.K.: Cambridge University Press, 1975.
- Récanati, F. *Direct Reference: From Language to Thought*. Oxford: Blackwell, 1993.
- Rorty, R. *Philosophy and the Mirror of Nature*. Princeton, NJ: Princeton University Press, 1979.
- Schwartz, S., ed. *Naming, Necessity, and Natural Kinds*. Ithaca, NY: Cornell University Press, 1977.

Timothy Williamson (1996)

## REFLECTIVE EQUILIBRIUM

Reflective equilibrium is a coherence method of philosophical justification or inquiry. Nelson Goodman (1955) introduced reflective equilibrium, although not under that name, to contemporary philosophy in a discussion of deductive and inductive logic. It is arguable, however, that philosophers have employed something such as reflective equilibrium to inquire into a wide range of topics since ancient times.

Goodman maintained that we justify an inference by showing that it conforms to the rules of either deduction or induction. But for the inferences to be justified, these rules must be valid. Goodman held that we justify rules of inference by showing that they accord with judgments we make about which particular inferences are acceptable and which are unacceptable. Goodman addressed the obvious objection to such a procedure as follows:

This looks flagrantly circular. I have said that deductive inferences are justified by their conformity to valid general rules, and that general

rules are justified by their conformity to valid inferences. But this circle is a virtuous one. A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend. The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement thus achieved lies the only justification needed for either.

(GOODMAN 1955, p. 67)

It is possible to read Goodman's proposal as purely methodological or as more epistemological. According to the epistemological reading, when we complete the process of mutual adjustment Goodman describes, thereby bringing our judgments regarding the particular inferences and the rules of inference we accept into a state of reflective equilibrium, these rules and particular judgments are by definition justified. According to this reading, being justified consists in being part of a system of beliefs, including rules and particular judgments, that has the sort of coherence that reflective equilibrium represents.

The methodological understanding of reflective equilibrium accepts that a philosophical inquiry into inductive or deductive inference is properly conducted by a process of mutual adjustment of the kind Goodman describes; it agrees that this is the best we can do in an attempt to justify the inferences we make and the rules of inference we accept. But the methodological reading is not definite about the epistemic status of the particular and general judgments we manage to bring into reflective equilibrium. It leaves open what positive epistemic status, if any, principles and judgments that are in reflective equilibrium might have. In addition, whereas the epistemological reading is committed to a coherentist account of justification, the methodological reading leaves open how best to account for the precise epistemic status (or statuses) attained by judgments that are in reflective equilibrium. Although Goodman probably intended an epistemological reading, let us adopt a methodological understanding of reflective equilibrium in order to keep as many epistemological options open as possible.

We owe the term reflective equilibrium to John Rawls (1971), who developed the method further and applied it to moral inquiry. If we think of the method as something to be applied in a stepwise fashion, which is useful even if not entirely accurate, then an inquirer, S, begins with a large set of initial moral judgments. These judgments will be intuitive for S in the sense that they are

cognitively spontaneous; they might concern propositions that are either particular or general. The first step on the road to reflective equilibrium, according to Rawls, is for S to eliminate certain initial moral judgments. For example, judgments that are not stable over time or in which S has little confidence should be dropped, as should judgments formed when S is emotionally distraught. In general, S eliminates those judgments formed in circumstances where there is some obvious reason for suspecting error. The remaining judgments will be S's considered moral judgments.

S's next task is to formulate a moral theory, that is, a set of moral principles that accounts for S's considered moral judgments. The coherence element of the method comes into play at this stage because S will grant neither considered moral judgments nor moral theory a privileged status. S will make revisions on both sides in the attempt to forge a coherent system of moral beliefs. When the emerging theory is found to conflict with central, very confidently made considered judgments, S must revise the theory. But if a well-confirmed element of the theory that is independently plausible is found to conflict with less firmly held considered judgments, then S will revise these judgments. S's decision regarding what to revise is made for each case on the basis of what seems most likely to be true or correct to S upon due consideration (there are various other ways of expressing this idea: We might, for example, say that S is to decide on the basis of S's degrees of belief or commitment or on the basis of what seems most plausible or acceptable to S).

To this point, Rawls's method corresponds with Goodman's, but Rawls does not allow S to stop here, at a point of narrow reflective equilibrium. According to Rawls, S must next consider alternatives to the moral theory that S accepts in narrow equilibrium along with philosophical arguments for and against S's own theory and the various alternatives S is considering. In his important work on reflective equilibrium, Norman Daniels (1979) argues that we can think of this as an attempt to attain coherence between the considered moral judgments and theory that S accepts in narrow reflective equilibrium and the background theories S accepts. The idea is that the philosophical arguments that S constructs will use premises drawn from among S's broader background beliefs, which might include such things as sociological views regarding the role of morality in society and philosophical or psychological theories regarding rational decision or the nature of persons.

An argument in favor of an alternative to the moral theory that S accepts in narrow reflective equilibrium that

is successful in the sense that S finds it compelling would, in effect, show that S's moral theory, considered moral judgments, and background beliefs are not coherent. As before, S is not bound to favor any type of belief when responding to such an argument; S must decide whether to revise considered moral judgments and moral theory or background beliefs on the basis of what seems most likely to be true to S after thorough reflection. When S attains a coherent system of considered moral judgments, moral theory, and background beliefs, S will have reached a state of wide reflective equilibrium.

An inquirer can certainly move from narrow to wide reflective equilibrium in the way that Daniels maintains, but this is not the only or most interesting way things can go. Consideration of alternative moral views and the relevant philosophical arguments can provide an occasion for a more radical type of revision of belief. Daniels seems to suppose that the only way in which considering alternatives to one's own view can force one to revise beliefs is by revealing that something else one believes, and believes more strongly, conflicts with one's prior view but coheres with the alternative. But it is quite clear that consideration of alternative moral and broader philosophical views can also lead one to revise beliefs in a way that is not dictated by one's prior beliefs and degrees of commitment. It is possible for an inquirer to find an alternative view attractive in its own right, even though it conflicts with everything the inquirer previously thought; and if upon reflection the inquirer finds the new alternative sufficiently attractive, he or she might well respond by accepting the alternative and revising his or her previous views.

Wide reflective equilibrium is best understood in a way that allows for this radical type of belief revision. According to this understanding, achieving wide reflective equilibrium is not simply a matter of rooting out conflicts among the beliefs one already holds and forging general principles that coherently account for one's considered moral judgments. It crucially involves exposing one's self to alternative moral and philosophical views with the knowledge that reflection upon such alternatives might lead one to make a radical break with one's previous views.

On this understanding, the ideal of wide reflective equilibrium is not defined merely as achieving coherence among all of one's beliefs: considered moral judgments, a moral theory, and background beliefs. The ideal crucially involves attaining a kind of reflective stability. Inquirers who have attained reflective equilibrium are, in effect, immune to threats from the inside and the outside. There will be no conflicts within such inquirers' systems of

belief, and in addition they can be confident that there are no alternatives to their own systems of belief that they would find more compelling than their own upon due reflection. The first sort of reflective stability is provided by reflective equilibrium on either understanding, the second only if reflective equilibrium is understood in a more radical way.

It is important to recognize that the essential feature of a belief revision that is radical, in the strict sense here at issue, is not the number or range of beliefs that are altered but rather the fact that the alteration is not continuous with the things that one previously believed. When a belief is revised in a way that is not strictly radical, the change is required in order to attain coherence among one's beliefs, and the alteration is dictated by other things that one believes more firmly than the belief that is revised. Particularly if the belief that is revised in this way concerns a general principle, the change can require revisions to a large number of other beliefs; such a revision would likely be called radical in common parlance, but it would not be radical in the strict sense.

When a change is one that counts as radical in the strict sense, the new beliefs come to seem compelling to one on their own, apart from their logical or evidential relations to one's previous beliefs. Indeed, the new belief will likely contradict things that one previously believed very strongly. Such a change may involve many beliefs or only a few. Philosophers, who might be guilty of considering such matters only abstractly, may find it difficult to accept the possibility of such radical changes in belief, but it is easy to find descriptions in novels, biographies, and autobiographies of people altering their views in ways that seem to be radical in the strict sense.

Because reflective equilibrium grants the inquirer's considered moral judgments a crucial role in inquiry, it has been widely criticized as a sophisticated version of intuitionism, making it an unreliable and extremely conservative method. Daniels (1979) sought to rebut the charge of intuitionism by arguing that reflective equilibrium is not compatible with foundationalism, which is a characteristic of intuitionism. Daniels's basic idea is that the inquirer's considered moral judgments do not function as intuitions because reflective equilibrium allows for such extensive revision of these judgments. This response seems to rely upon too narrow a conception of foundationalism; in particular, it seems to suppose that the beliefs that serve the foundational role must be identifiable in advance of inquiry and also be unrevisable, or at least relatively unrevisable.



Given these suppositions it is natural to think that if reflective equilibrium is a foundationalist method, the inquirer's initial considered moral judgments must be the foundations, and then conclude that it cannot be a foundationalist method because these judgments are subject to way too much revision. However, if one supposes instead that the foundations might emerge through a course of inquiry, it is unclear that reflective equilibrium does not constitute a version of foundationalism. During a person's inquiry there will be various relatively strongly held judgments that determine the course of the inquiry and the views the person comes to hold in reflective equilibrium. It is unlikely that all these judgments will be drawn from among the inquirer's considered moral judgments, but almost certainly many will. Others might come from among the person's background beliefs, some might concern moral principles, and perhaps some will be about which member of a conflicting set of beliefs should be revised.

But that does not really matter. The fact remains that at the end of inquiry, it will be possible to identify a set of judgments that provide a psychological basis for the rest of the beliefs the person holds in reflective equilibrium. Many, although perhaps not all, of these judgments will probably be intuitive in the sense that they are cognitively spontaneous. It is possible, therefore, that these intuitive judgments serve as an epistemological foundation for the rest of the beliefs the person holds in reflective equilibrium. It remains to be seen whether this constitutes a ground for objection to reflective equilibrium.

Reflective equilibrium fares better when it comes to the charge that it is extremely conservative. The extensive revisability of considered moral judgments may show that it can be construed as a version of intuitionism, but it surely shows that the method is not guaranteed to produce nothing more than a cleaned up, systematized version of conventional morality. The method has the potential to, and indeed is likely to, lead many inquirers to make extensive changes to their moral views. A particular inquirer might, of course, end up holding very conventional views in reflective equilibrium. Indeed, this is just what will happen if the inquirer is more strongly committed to enough elements of conventional morality than he or she is to anything that conflicts with them, and retains these commitments through the course of reflection upon alternatives to and criticisms of conventional morality. But it is not clear that a method of moral inquiry is inadequate unless it absolutely excludes this possibility.

More worrisome is a general fact illustrated by the possibility just considered: Given the right (or perhaps one should say wrong!) moral judgments and background beliefs held strongly and tenaciously enough, it would seem to be possible for an inquirer to end up holding virtually any moral view, even a bizarre or repugnant view, in reflective equilibrium. The worry is not confined to reflective equilibrium when used as a method of moral inquiry. No matter what a person might use this method to inquire about, given the right intuitive beliefs held with sufficient strength and tenacity, the person could end up holding virtually any view one could imagine in reflective equilibrium: extreme skepticism, solipsism, nihilism, anarchism, totalitarianism, atheism, or theism—you pick whatever views you think are beyond the pale. How then could anyone take reflective equilibrium to be an acceptable approach to moral inquiry, or philosophical inquiry more generally? As various critics have put the point: The method clearly leads an inquirer to the coherent position he or she finds most acceptable, the position that best preserves beliefs to which he or she is most strongly committed, but why think this position is anything more than that, in particular, why think it is true or likely to be true?

One might have once thought that this fundamental objection is really pressing only against reflective equilibrium when used to inquire into morality and other such things, where it is all too obvious that different people can hold, and hold very strongly, very different and incompatible considered judgments. One might have presumed that for such purposes as working out valid rules for deductive or inductive inferences, which is what Goodman originally proposed that the method be used for, there is no real problem because there just is not the same sort of diversity and conflict between the strongly held considered judgments of different people. But as Stephen Stich most particularly has stressed, empirical work has shown that many of the inferences ordinary people find intuitively acceptable are in fact fallacious. Hence, we can foresee that the rules of inference these people accept in reflective equilibrium will not be valid. So we cannot even trust reflective equilibrium to be an acceptable method of inquiry for those areas where it was originally proposed.

It might seem, therefore, that Rawls's early critics were right to argue that unless we can find some reason for trusting the reliability of the intuitive judgments that play such a crucial role in the method of reflective equilibrium, this method of inquiry cannot be acceptable. Daniels (1979) was perhaps right when he claimed, in response, that it is unreasonable to expect such a reason to be provided before we begin our inquiries and that

such a reason should only be expected to emerge as part of the overall system of beliefs accepted in reflective equilibrium. But surely it seems unreasonable to hope that it will ever be possible to offer even such an internal defense of reflective equilibrium as a method that is reliable for all who might employ it. Such a defense would require that all, or nearly all, inquirers employing the method converge on the same theory, and this seems rather unlikely. In addition, this sort of defense seems to underestimate the obstacles facing reflective equilibrium: It is not just that there is no reason to think the method is reliable, but also none for suspecting that it is unreliable—there are fairly strong reasons for believing that the method is unreliable, that it is not the case that a very high proportion of those who employ it will be led to accept a system of beliefs that is largely correct.

Nevertheless, it is possible to offer a defense, albeit a modest defense, of reflective equilibrium. The first step is to recognize that there are a number of different positive epistemic statuses. For simplicity, let's distinguish only two. The first is the positive epistemic status that plays the major role in distinguishing knowledge from mere true belief. This status is most commonly referred to as justification or warrant. Attempts to account for justification in terms of the reliable formation of belief have been popular and influential. Even if such attempts fail, the majority of epistemologists would still maintain that there is some sort of strong connection between justification and truth: justified beliefs must, in some sense, be likely to be true. The second positive epistemic status is the sort of subjective rationality that Richard Foley has stressed. A belief is rational in this sense when it satisfies the believer's own epistemic standards, that is, when the believer would consider the belief likely to be true after due reflection. Unlike justification, there seems no reason to suppose that beliefs that are rational in this sense are likely to be true.

Having distinguished these two positive epistemic statuses, it should be fairly clear that reflective equilibrium can, in fact, be guaranteed to lead the inquirer to hold rational beliefs. It should also be easy to see that an inquirer who deviated from reflective equilibrium would be led to hold some beliefs that are not rational because, in order to deviate, the inquirer would have to resolve some conflict by rejecting a belief that, upon reflection, the inquirer considers more likely to be true than the belief being retained. What reflective equilibrium cannot guarantee every inquirer is justified beliefs. If it followed that by employing the method of reflective equilibrium an inquirer was sure to form rational beliefs but equally

sure to form unjustified beliefs, and hence fail to attain knowledge, the method would indeed be unacceptable. But it would be hasty to infer that no inquirer employing reflective equilibrium can be led to hold justified beliefs simply because the method cannot guarantee justified beliefs to all inquirers.

In an influential paper on the method of moral inquiry written years before he advocated reflective equilibrium, Rawls (1951) argued that we should construct a moral theory by formulating principles that account for the considered moral judgments of competent moral judges. Whereas the notion of considered moral judgments is used in describing reflective equilibrium, the notion of the competent moral judge has fallen by the wayside. But suppose, as we clearly do in our ordinary lives, that some people are competent moral judges, whereas others are not. We ordinarily suppose that the intuitive moral judgments of competent moral judges are reliable. We might be wrong, and of course we might also be right, that those people who have the characteristics we commonly associate with moral competence, in fact, make a person a reliable moral judge. So let us understand competent moral judges as those whose intuitive moral judgments are reliable. The beliefs competent moral judges would hold in reflective equilibrium obviously would be reliable.

If one condition for being justified is that a person must not only be reliable, but be able to prove that he or she is, then perhaps even the beliefs held by competent moral judges in reflective equilibrium are not justified. For competent judges will not be able to prove to incompetent judges that they are reliable. But this condition for being justified is almost certainly too strong: If we were to apply it across the board, we would know little or nothing. If what is necessary for justification is only that one is reliable, not that one be able to prove that one is, then the beliefs competent moral judges hold in reflective equilibrium may, for all that has been said so far, be justified.

This is not, of course, all that we might have wanted in the way of a defense of reflective equilibrium. We cannot prove that anyone is or is not a competent judge. We cannot prove which characteristics make for competent judges and which make for incompetent judges. Perhaps there are no competent judges. Perhaps there are, but even they do not know that they are. But it is not outlandish to think that there are competent moral judges and that most of us know something about who they are and what they are like. But for the sake of this argument, suppose only that it is possible that there are competent

moral judges. If there are any competent judges, then the beliefs they hold in reflective equilibrium are justified. And this suggests that these beliefs, or at least many of them, might count as knowledge.

So we can say this much about reflective equilibrium. It is the only rational method of inquiry and it is possible that, by employing this method, a person will be led to hold justified beliefs and to attain knowledge. This is certainly less than one would like to be able to say in support of a method of philosophical inquiry, but it is sufficient to show that reflective equilibrium is an acceptable method for ethics, and philosophy more generally.

**See also** Applied Ethics; Goodman, Nelson; Logic, History of; Modern Logic: From Frege to Gödel; Metaethics; Moral Epistemology; Rawls, John.

### Bibliography

- Bonevac, Daniel. "Reflection without Equilibrium." *The Journal of Philosophy* (2004).
- Brandt, Richard. *A Theory of The Good and The Right*. Oxford: Clarendon Press, 1979.
- Brink, David O. *Moral Realism and The Foundations of Ethics*. Cambridge, U.K.: Cambridge University Press, 1989.
- Daniels, Norman. *Justice and Justification: Reflective Equilibrium in Theory and Practice*. New York: Cambridge University Press, 1996.
- Daniels, Norman. "On Some Methods of Ethics and Linguistics." *Philosophical Studies* 37 (1980): 21–36.
- Daniels, Norman. "Wide Reflective Equilibrium and Theory Acceptance in Ethics." *Journal of Philosophy* 76 (1979): 256–282.
- DePaul, Michael R. *Balance and Refinement: Beyond Coherentism in Moral Inquiry*. London: Routledge, 1993.
- DePaul, Michael R., and William Ramsey. *Rethinking Intuition: The Psychology of Intuitions and Their Role in Philosophical Inquiry*. Lanham, MD: Rowman and Littlefield, 1998.
- Foley, Richard. *The Theory of Epistemic Rationality*. Cambridge, MA: Harvard University Press, 1987.
- Goodman, Nelson. *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press, 1955.
- Hare, R.M. "Rawls' Theory of Justice—I." *The Philosophical Quarterly* 23 (1973).
- Harman, Gilbert. "Moral Philosophy, and Linguistics." In *Proceedings of The Twentieth World Congress of Philosophy*. Vol. I: *Ethics*, edited by Klaus Brinkmann. Bowling Green, Ohio: Philosophy Documentation Center, 1999.
- Harman, Gilbert. *The Nature of Morality*. New York: Oxford University Press, 1977.
- McNaughton, David A. *Moral Vision*. Oxford: Oxford University Press, 1988.
- Nussbaum, Martha. *Love's Knowledge*. Oxford: Oxford University Press, 1990.
- Rawls, John. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1971.
- Rawls, John. "Outline of a Decision Procedure for Ethics." *The Philosophical Review* 60 (1951): 177–197.
- Sencerz, Stefan. "Moral Intuitions and Justification in Ethics." *Philosophical Studies* 50 (1986): 77–95.
- Singer, Peter. "Sidgwick and Reflective Equilibrium." *Monist* 58 (1974): 490–517.
- Sosa, Ernest. "Equilibrium in Coherence?" In *Knowledge in Perspective: Selected Essays in Epistemology*. Cambridge, U.K.: Cambridge University Press, 1991.
- Stich, Steven. *The Fragmentation of Reason*. Cambridge, MA: MIT Press, 1990.
- Stich, Stephen, and Richard Nisbett. "Justification and The Psychology of Human Reasoning" *Philosophy of Science* 47 (1980): 188–202.
- Timmons, Mark. "Foundationalism and The Structure of Ethical Justification." *Ethics* 97 (1987): 595–609.
- Weinberg, Jonathan, Shaun Nichols, and Stephen Stich. "Normativity and Epistemic Intuitions." *Philosophical Topics* 29 (2001): 429–460.

Michael R. DePaul (2005)

## REFORMATION

In the narrower and probably most common sense, "Reformation" is the name given to the spiritual crisis of the sixteenth century that resulted in the permanent division of the Western church. The birthdate of the Reformation is traditionally given as 1517, the year in which Martin Luther posted his Ninety-five Theses on the door of the Castle Church in Wittenberg; the termination of the period may be assigned to the 1550s, by which time an ecclesiastical stalemate between the Protestants and the Roman Catholics appeared unavoidable. Sometimes the Reformation is extended backward to include such early reform movements as Lollardy or forward to include the religious conflicts, lasting into the seventeenth century, that sought to resolve the Catholic-Protestant stalemate forcibly or to readjust the divisions between the various Protestant groups. Reformation describes the aspirations of the age rather than its achievements. The Protestants did not succeed in reforming the church but only in splitting it into rival groups, each of which claimed for itself the fulfillment of the old dream of reformation in head and members.

### THE AGE OF REFORMATION

The Protestant movement was not the only attempt to bring the dream into reality. It can, indeed, be correctly interpreted only in relation to other reform movements even if we determine not to include these under the same general descriptive label. The sixteenth century was the age of reformation (or of reformations, in the plural), not

just of the Reformation, and this is a fact of some importance in assessing the impact of the spiritual crisis on Western intellectual history. We should distinguish four reform groups in the sixteenth century, each of which left its own distinctive mark on Western culture.

**HUMANISTIC REFORMERS.** The humanists were not merely (as Luther himself thought) forerunners who prepared the way for the Protestants. They developed a reform program of their own that did not lead to the formation of independent institutions but continued, even after the appearance of Luther, to exercise influence from within both of the two main confessional groups. The foremost humanistic reformer in northern Europe was Desiderius Erasmus, who wished to purify the church by returning to its primitive sources—the New Testament and the writings of the Fathers. His “philosophy of Christ” minimized the dogmatic and the institutional and treated Christ mainly as a teacher of virtue and Christianity as an ethical affair not essentially different from the pagan philosophies. Although not less critical of ecclesiastical abuses than was Luther, Erasmus deplored any action that might disrupt the unity and peace of Christendom, and this was one of the reasons that he remained aloof from the Protestant Reformation.

**RADICAL REFORMERS.** “Radical reformers” is a general term for a variety of groups and individuals who felt that the Protestant leaders had not gone far enough and that reform could not be brought about without abandoning the old idea of the state church (the *corpus Christianum*). Of these radical or left-wing reformers, the Anabaptists (Swiss Brethren, Hutterites, and Mennonites) were biblical literalists who sought to establish voluntary associations of the regenerate on the New Testament pattern. The spiritualists (Andreas Carlstadt, Thomas Münzer, Sebastian Franck, Caspar Schwenckfeld), appealing to the Spirit who caused the Scriptures to be written, laid claim to immediate converse with God. The rationalists (notably the two Socinus) read the Bible in the light of reason even when reason led them to deny Christ’s full deity and atoning sacrifice. A few of the radicals (for example, the leaders of the Münster uprising in 1534) were revolutionaries who brought total destruction upon themselves; many, like Michael Servetus, were free spirits who founded no school, but the influence of others, despite brutal persecution by Roman Catholics and Protestants alike, still survives in some present-day denominations and sects.

**CATHOLIC REFORMERS.** The Roman Catholics rejected the Protestant reform as essentially a revolt against the church, and they sought renewal of the church by the twofold means of fostering a churchly piety and taking an official stand on the administrative and dogmatic demands of the “heretics.” Two of the greatest landmarks of the Catholic reformation were the establishment of the Jesuit order under the leadership of Ignatius Loyola and the work of the Council of Trent (1545–1563). The council, not without political and theological difficulties, sought to repudiate Protestant errors on authority, justification, and the sacraments. Yet the Tridentine fathers opposed many of the practical abuses and even theological inadequacies that had first provoked the Protestant movement. Preoccupation with Protestant errors, together with the militant campaign of suppression that followed the council, make it not inappropriate to speak of the Catholic reformation as the Counter-Reformation, though it was not merely this and had its roots in pre-Lutheran piety.

**THE PROTESTANT REFORMATION.** The Protestant leaders (the reformers in the narrower sense) were themselves not strictly a single group. Protestantism took three distinctive, though fundamentally related, forms.

Lutheranism, rooted in the religious struggles of Luther and his revolt against the papacy, prevailed in most of Germany and was wholly victorious in the Scandinavian countries. It was the Lutheran princes and cities represented at the Imperial Diet of Speyer in 1529 who, by making their historic protest, gave the Lutheran movement its nickname Protestantism. The classic formulation of Lutheran belief is the Augsburg Confession of 1530.

The so-called Reformed churches grew up first in Switzerland (under Huldrych Zwingli and John Calvin); won majorities in Scotland, Holland, and parts of Germany; and maintained strong pockets of influence in France, England (where they were called Presbyterians), and eastern Europe. From their beginning they were a less homogeneous group than the Lutherans and produced a variety of national confessions rather than a single statement comparable to the Lutheran Augsburg Confession. Nevertheless, the Lutheran interpretation of the Gospel exercised a decisive influence over the Reformed confessions, and though Zwingli sought to affirm his relative independence from the Germans, Calvin was one of Luther’s staunchest admirers.

The Anglican reformation proceeded slowly, largely for political reasons. The repudiation of papal authority by Henry VIII, though not intended to alter Catholic doc-

trine, left the door open to Protestant reform in the reign of his son Edward VI, and the Romanizing reaction under Mary only temporarily reversed the trend. The Thirty-nine Articles of Religion (Latin 1563, English 1571), adopted under Elizabeth I as the official doctrinal standard of the reformed Church of England, are largely a compilation of Continental Protestant ideas. Parts of the Lutheran confessions of Augsburg and Württemberg are reproduced verbatim, and the articles on predestination and the Eucharist are clearly indebted to Reformed (Calvinistic) theology.

### ESSENTIAL PROTESTANT DOCTRINES

In all three of its branches the Protestant Reformation was inextricably bound up with social and political factors, so that its triumph was always, in the final analysis, contingent on governmental support. Nevertheless, it was essentially a religious movement and its theological ideas have left their mark on European intellectual history—sometimes, however, because they have been misinterpreted or interpreted too one-sidedly. Three beliefs are particularly associated with the Protestant movement: the authority of the Word, justification by faith alone, and the priesthood of all believers. These beliefs have frequently been explained as the advent of individualism in the religious sphere, as though the intention were to regard the individual as his own priest with immediate access to God, to leave him in solitude with his conscience and his Bible, or to make each man his own pope in the interpretation of Scripture.

Fundamentally, however, the original Protestant reformers were suspicious of “immediate access to God,” which they associated with the spiritualists, and they sought, rather, to replace the medieval notion of institutional means with a concept of the Christian fellowship as the locus of God’s Word. The Word of God was understood chiefly as an effective proclamation of the Gospel, based on the Scriptures, which evokes faith and sustains a fellowship of believers each of whom is priest to his brothers. The heart of this proclamation is the promise of free forgiveness (justification) through Christ, which needs only to be accepted by the faith that is awakened through the proclamation itself. We may perhaps add a fourth idea of great religious and even social consequence: vocation—that the good works required of the justified man are not so much special religious acts as the thankful performance of his calling for the good of his neighbor. These four ideas were held in common by all three Protestant groups, and their formulation may be traced to Luther himself. Characteristic differences

among the groups also developed; for example, the Reformed differed from the Lutherans, as is well known, on the manner in which the benefits of Christ’s Passion are received in the Eucharist.

### THE REFORMATION AND WESTERN THOUGHT

The Reformation’s role in the making of the modern mind is a complex question that has ramifications in areas as diverse as social, economic, political, and artistic history as well as in the history of philosophy and science. Sometimes the Reformation has been represented as the great watershed between the medieval and modern worlds. This is, perhaps, partly because the individualism of Reformation thought has been overestimated and partly because certain isolated events in Luther’s life—the burning of the papal bull, the defiant stand before the Diet at Worms—have deeply impressed themselves on the German imagination. In some respects, however, the Reformation can be better understood as a late phase of medieval history than as an early stirring of the modern mind. The fundamental concerns of Luther were medieval, and it may be argued that in giving fresh vitality to religious questions he merely postponed for a while the triumph of Renaissance secularism. Moreover, though the Protestant reformers spoke ideally of a communion of saints (believers), in practice they refused to abandon the medieval concept of a Christian society (that is, an authoritarian, church-dominated society).

Unquestionably, the very existence of the Protestant churches alongside the Roman Catholic Church weakened the authoritarian ideal. But this was an accidental product of the Reformation—a consequence, indeed, of its failure rather than of its cherished principles. It was the humanistic reformers, not the Protestants, who undermined the dogmatic conception of religion, and it was the radicals who broke with the old alliance between the spiritual and secular arms of the *corpus Christianum*. Similarly, if, as has been argued, Calvinistic ideas had revolutionary economic and political consequences, this was hardly the reformer’s intention. On the other hand, the Reformation did, by its very nature, make a powerful impact on literature and music, education and scholarship; even its influence on the visual arts was not always uncreative.

### REFORMATION AND SCIENCE

The chief contribution of the Reformation to the history of Western philosophy was no doubt the accidental one of helping philosophy toward autonomy by weakening

ecclesiastical domination. Attempts to establish the influence of Lutheran ideas on some of the German philosophers are often interesting but seldom of very great importance and sometimes farfetched. It might have been a service for philosophy, as it was for theology, that Luther shattered the medieval synthesis of Christianity and Aristotelianism, but the reformer's immediate successors reinstated the Greek philosopher, and the Christian faith was perilously entangled in an obsolete cosmology. (Ironically, Philipp Melanchthon repudiated Copernican astronomy on the ground that it represented merely a revival of outmoded theories that had already been rejected in the ancient world.)

Luther himself prepared the way for the conflict of theology and the modern worldview by refuting a scientific theory on theological grounds—if, indeed, the notorious passage from the *Table Talk*, “Joshua commanded the sun, not the earth, to stand still,” is authentic. Yet an open clash of science and religion was not unavoidable until post-Reformation theologians in the age of Protestant scholasticism had reaffirmed the old partnership with Aristotelianism and had come to think of the Scriptures as containing a “biblical science” that could compete with Copernican science. Luther and Calvin themselves did not accept the Ptolemaic cosmology in defiance of scientific evidence since the weight of the evidence during their lifetimes was still against Nicolas Copernicus. In principle, they were not suspicious of scientific progress. On the contrary, Luther welcomed the stirring of the new science, in which he saw a partial recovery of Adam's lost dominion over nature, and Calvin envied the astronomer's closeness to the mind of the Maker. They were both interested in the Bible not as an encyclopedia of supernaturally communicated information but as the vehicle of Christ's presence to his church in the Gospel proclamation.

Luther had grasped clearly that theological and scientific interest in nature are two distinct things. For example, from the religious viewpoint the light of the moon was for him a symbol of divine care, but he recognized that the astronomer's concern was to show how the moon's light was borrowed from the sun. Similarly, Calvin argued that biblical observations on the heavenly bodies, such as those in Genesis and the Nineteenth Psalm, are not scientific statements but homely forms of speech accommodated to the unlearned. Luther understood even better than Calvin that theology's heaven is not the same as the astronomical heavens; hence, the celebrated *Dextera Dei est ubique* (God's right hand is everywhere). Elementary though they may seem today, such

concessions and insights, had they not been neglected or expressly repudiated by Protestant orthodoxy, could have saved the Reformation churches from their warfare with science. Conversely, they might have prevented skeptics from drawing overhasty theological conclusions from natural science.

**See also** Aristotelianism; Calvin, John; Copernicus, Nicolas; Erasmus, Desiderius; Franck, Sebastian; Humanism; Luther, Martin; Melanchthon, Philipp; Servetus, Michael; Socinianism.

### Bibliography

- Studies of the Reformation include Roland H. Bainton, *The Reformation of the Sixteenth Century* (Boston: Beacon Press, 1952); Harold J. Grimm, *The Reformation Era* (New York: Macmillan, 1954); G. R. Elton, ed., *The New Cambridge Modern History*, Vol. II, *The Reformation 1520–1559* (Cambridge, U.K.: Cambridge University Press, 1958); Émile G. Léonard, *Histoire générale du protestantisme*, Vol. I, *La réformation* (Paris, 1961); H. A. Enno van Gelder, *The Two Reformations in the 16th Century: A Study of the Religious Aspects and Consequences of Renaissance and Humanism* (The Hague, 1961); George Huntston Williams, *The Radical Reformation* (Philadelphia: Westminster Press, 1962); and H. Daniel-Rops, *The Catholic Reformation*, translated by John Warrington (New York: Dent, 1962).
- The Augsburg Confession, the Thirty-nine Articles, and a representative Reformed confession will be found in *The Faith of Christendom: A Source Book of Creeds and Confessions*, edited by B. A. Gerrish (New York, 1963). For the religious ideas of the reformers there are excellent bibliographies in the studies by Grimm and Léonard.
- On the special question of the Reformation in relation to Western culture and thought, see Heinrich Bornkamm, *Luther's World of Thought*, translated by Martin H. Bertram (St. Louis: Concordia Publishing House, 1958); John Dillenberger, *Protestant Thought and Natural Science: An Historical Interpretation* (Garden City, NY: Doubleday, 1960); Karl Holl, *The Cultural Significance of the Reformation*, translated by Karl Hertz and Barbara Hertz and John H. Lichtblau (New York: Meridian, 1959); Jaroslav Pelikan, *From Luther to Kierkegaard: A Study in the History of Theology* (St. Louis: Concordia Publishing House, 1950); and Ernst Troeltsch, *Protestantism and Progress: A Historical Study of the Relation of Protestantism to the Modern World*, translated by W. Montgomery (London: Williams and Norgate, 1912).
- OTHER RECOMMENDED WORKS**
- Brady, Thomas A., Jr. *The Protestant Reformation in German History*. Washington, DC: German Historical Institute, 1998.
- Caponetto, Salvatore. *The Protestant Reformation in Sixteenth-Century Italy*. Translated by Anne C. Tedeschi and John Tedeschi. Kirksville, MO: Thomas Jefferson University Press, 1999.
- Gonzalez, Justo L. *A History of Christian Thought*. Nashville: Abingdon Press, 1970–1975.

- Gray, Madeleine. *The Protestant Reformation: Belief, Practice, and Tradition*. Brighton: Sussex Academic Press, 2003.
- Harrison, Peter. "Original Sin and the Problem of Knowledge in Early Modern Europe." *Journal of the History of Ideas* 63 (2) (2002): 239–259.
- Hillerbrand, Hans Joachim, ed. *The Protestant Reformation*. New York: Walker, 1968.
- Johnston, Andrew. *The Protestant Reformation in Europe*. New York: Longman, 1991.
- Kourvetaris, George A. *Social Thought*. Lanham, MD: University Press of America, 1994.
- Painter, Mark A. *The Depravity of Wisdom: The Protestant Reformation and The Disengagement of Knowledge from Virtue in Modern Philosophy*. Brookfield, VT: Ashgate, 1999.
- Spitz, Lewis William, ed. *The Protestant Reformation*. Englewood Cliffs, NJ: Prentice-Hall, 1966.
- Spitz, Lewis William. *The Protestant Reformation, 1517–1559*. New York: Harper & Row, 1985.
- Steinmetz, David Curtis. *Luther and Staupitz: An Essay in the Intellectual Origins of the Protestant Reformation*. Durham, NC: Duke University Press, 1980.
- Thompson, Stephen P., ed. *The Reformation*. San Diego: Greenhaven Press, 1999.

**B. A. Gerrish (1967)**

*Bibliography updated by Tamra Frei (2005)*

## RÉGIS, PIERRE-SYLVAIN (1632–1707)

Pierre-Sylvain Régis was a student of the Cartesian physicist Jacques Rohault. Like Rohault, Régis expounded Cartesianism in public lectures. In 1680 François de Harlay de Champvallon, the archbishop of Paris, told Régis that King Louis XIV forbade public lectures for fear of uproar concerning the Cartesian explanation of transubstantiation. Régis continued to give private lessons, and by 1699 the conflict over Cartesianism had subsided, leading to his admission to the Académie des Sciences, along with Nicolas Malebranche, whose occasionalist philosophy was a response to problems of Cartesian dualism.

Régis's system is based on fourteen self-evident metaphysical principles derived from the *cogito* (Descartes' basic axiom: "I think, therefore I am").

- (1) All properties belong to something, that is, nothing can have no properties
- (2) All effects presuppose causes
- (3) An effect can have no more perfection than does its total cause
- (4) All changes in a subject proceed from an external cause

- (5) All modes presuppose substances in which to exist
- (6) A mode that modifies one kind of substance cannot modify any other kind of substance
- (7) All that exists is either a substance or a mode
- (8) Essences are indivisible; if anything is added to or subtracted from an essence, it is destroyed
- (9) Privations and negations are known only by their opposites
- (10) External things are known only by way of ideas; what one has no idea of is to one as though nonexistent
- (11) All ideas, to the extent that they represent properties, depend on their objects as exemplary causes
- (12) The exemplary cause of an idea contains formally all the perfections the idea represents
- (13) Facts attested by many people of diverse times, nations, and interests as known in themselves and of which one cannot suspect conspiracy to support a lie should be accepted as constant and indubitable as though one had viewed them oneself
- (14) Witness of infinitely powerful, wise, good, and truthful God should have as much persuasive force on one's mind as the most convincing reasons

The ontological principles are to the effect that whatever exists is either a substance or a modification of a substance and that modifications cannot belong to nothing. The basic epistemological principle is that external material bodies can be known only by way of representational ideas that are themselves mental modifications of the mind. The central principle is that all effects presuppose causes that must have as much or more perfection than their effects. All modifications are effects and thus are ultimately caused by a substance. All ideas (both sensory images and intelligible concepts) are also effects caused by some substance, and all of them represent the perfections of their causes. The basic problem with these assertions is that mind whose essence is active unextended thinking is essentially unlike matter whose essence is passive unthinking extension. Thus, all Cartesians must face the two questions first posed to René Descartes by Princess Elisabeth of Bohemia: Given that Cartesian mind and matter are essentially unlike one another, how can they interact causally? And how can a mind know extended matter by way of unextended sensations and ideas? Régis answers as Descartes does that one knows causal interaction takes place and God can make this hap-

pen, even if one does not understand how. The question of how mental ideas can represent material bodies that are essentially unlike mental things is also settled in the Cartesian way with the assertion that God makes it so, even if one does not understand how.

Régis maintains that his metaphysics, logic (which follows Antoine Arnauld's), and ethics (based on self-interest) are certain and complete. Physical explanations are also based on self-evident principles, but those humans give are only probable because, on the principles of Cartesian physics, several explanations are deducible for each event, and one does not know which one God chose. Régis says that the simplest is the most probable. If one had complete knowledge of the Cartesian deductive system as God does, however, one would have certain knowledge in physics, and one should keep that goal in mind. But the search must be made systematically. Régis, like Descartes, opposes "arbitrary hypotheses," explanations not deducible from self-evident principles within a system. Ad hoc explanations that are not part of a comprehensive theory are useless. Régis, like Descartes and all later Cartesians, believes the correct theory is mechanism.

The most distinctive feature of Régis's Cartesianism is his doctrine that man is a compound substance. In this union, eight conditions pertain:

- (1) The soul always has the idea of extension
- (2) Specific brain movements cause specific ideas
- (3) Animal spirits cause motions in the brain that give rise to imagination, sensation, and memory of material objects
- (4) Pleasure and pain are signs of bodies suitable and unsuitable to the human body
- (5) Man has a penchant to love or hate, and pursue or flee, the objects of pleasurable or painful ideas
- (6) Sentiments and passions lead to actions of the body toward self-preservation
- (7) The soul thinks of particular bodies only when particular brain movements occur
- (8) The union holds only so long as the body is alive and functions properly

Because ideas must have existing exemplary causes, Régis argues contrary to Descartes that one knows both the essence and existence of both mind and matter.

Man, Régis explains (as did Descartes' Dutch disciple Regius [Henry de Roy]), is an accidental union of mind or soul and body. Descartes himself adamantly opposes

this view by insisting that the union is substantial, not accidental. If the union is only accidental, this makes the mind or soul a property of the body, not a substance on equal standing with the material substance. Then when the body dies, this accidental mind or soul would disintegrate with the other bodily properties. Régis argues that in fact the soul disintegrates and the mind survives, but it cannot think temporally—because this depends on bodily motions—and instead can contemplate only itself and God.

Like Descartes, Régis shows how operations of the body take place through actions of external bodies on sense organs to cause movements in the brain. He admits that causal interaction between mind and body is inexplicable and can be accepted only on faith. It is a brute fact that because of its union with a body, a mind or soul has the idea of extension and can cause that body to move. And because of the union, distinctive brain movements always give rise in the mind to distinctive sensations and concepts of the material objects affecting the brain.

Descartes asserts that one is born with innate ideas of mind, God, and matter. But Régis says that all ideas, even of God, depend on brain movements. Thus for Régis, after separation of mind and body at death, the mind no longer has the idea of extension, and no imagination or memory of, or power over, the material world.

Régis claims that Pierre-Daniel Huet thinks Descartes is a skeptic because Huet does not distinguish methodological from real doubt. Also, Huet is wrong to argue that Descartes' explanation of transubstantiation does not preserve the body of Christ in the sacrament and to claim that because Descartes believes God has the power to do anything, one has no certain knowledge of one's world. Régis shows that Malebranche's theory of seeing all things in God requires an impossible union of man with God. Jean Du Hamel is accused of failing to see that mental ideas that do not resemble their material objects still make these objects known. Régis insists that Benedict (Baruch) de Spinoza fails to see that God is not an ordinary substance and thus confuses the material world with God.

Like Rohault, Régis insists that reason and faith do not conflict. Reason is infallible in the order of nature; faith, in the order of grace. Events in one order cannot be explained with principles of the other. Thus, Régis offers no physical explanation of transubstantiation—as do Descartes, other Cartesians, and the scholastic physicists. He argues that transubstantiation is an event not in the order of nature but in the order of grace.



**See also** Arnauld, Antoine; Cartesianism; Descartes, René; Elisabeth, Princess of Bohemia; Essence and Existence; Huet, Pierre-Daniel; Malebranche, Nicolas; Regius, Henricus (Henry de Roy); Rohault, Jacques; Spinoza, Benedict (Baruch) de.

### **Bibliography**

Schmaltz, Tad M. *Radical Cartesianism: The French Reception of Descartes*. New York: Cambridge University Press, 2002.

Watson, Richard A. *The Breakdown of Cartesian Metaphysics*. Indianapolis: Hackett, 1998. Originally published as *The Downfall of Cartesianism, 1673–1712: A Study of Epistemological Issues in Late Seventeenth-Century Cartesianism* (The Hague: Nijhoff, 1965).

**Richard A. Watson (1967, 2005)**

## REGIUS, HENRICUS (HENRI DE ROY) (1598–1679)

Regius (Henri de Roy), a Dutch academic, was a major figure in disputes over Cartesianism in Utrecht, The Netherlands, during the seventeenth century. Regius received a medical degree from the University of Padua before returning in 1638 to his hometown of Utrecht to become a professor of medicine and botany at the university. Before his appointment there he gave private lectures based on the *Dioptrics* and *Meteors*, two of the essays published with René Descartes's *Discourse on the Method* (1637). What Regius found to be particularly congenial in these texts was the proposal there that observable phenomena be explained in terms of the mechanical properties of insensible material parts.

In 1641 Regius took advantage of his good relations with the new rector of the university, Gisbertius Voetius, to obtain permission to submit for discussion various “medical disputations.” The first two disputations provide a mechanistic reinterpretation of Aristotelian notions, but in the third disputation Regius took the more aggressive tack of claiming that the union of the soul and body is not substantial, as the Aristotelians claimed, but accidental. Voetius responded with an appendix that defended Aristotelianism against the “new philosophy,” and on the advice of Descartes, Regius offered a response that suggested that the Aristotelians had difficulty avoiding atheism. In 1642 the burgomasters of Utrecht ordered the confiscation of Regius's response and endorsed a statement by the faculty that condemned the teaching of the new philosophy. Descartes intervened

by publishing attacks on Voetius in 1642 and 1644 that the burgomasters judged to be libelous. Fearful of imprisonment, Descartes sought the protection of the French ambassador, who succeeded in suppressing his arrest warrant.

To this point, Descartes had a favorable opinion of Regius. However, matters took a turn for the worse in 1645 when Regius sent Descartes a draft of his *Fundamenta physices*. Descartes was shocked by the assertion in one section of this text that it is impossible to prove that the soul is anything more than a mode of body. When Regius went ahead and published his text in 1646, Descartes denounced it in the preface to the French edition of the *Principles* (1647). A student of Regius published a broadsheet that highlighted Regius's rejection not only of a proof of immortality but also of innate ideas and the possibility of a proof of the existence of the material world. Descartes responded in 1648 with his *Notes against a Broadsheet*, and Regius replied that same year with his *Brevis explicatio mentis humanae*. Regius's text included a letter from Petrus Wassenaer defending Regius against the charge in the preface to the *Principles* that he had plagiarized portions of Descartes's unpublished treatise on animals. After Descartes's death Regius published second and third editions (1654 and 1661, respectively) of the *Fundamenta physices* with the new title *Philosophia naturalis*, in which he attempted to further defend the project of freeing mechanistic physics and physiology from dogmatic metaphysics, on which Descartes had attempted to found it. Regius's presentation there of the new science as a system of probable hypotheses is similar to the one found in the *Traité de physique* (1671) of the French physicist Jacques Rohault, perhaps the most influential defense of Cartesian physics in the century following Descartes's death.

**See also** Aristotelianism; Cartesianism; Descartes, René; Philosophy of Physics; Rohault, Jacques.

### **Bibliography**

Dijksterhuis, E. J., et al. *Descartes et le cartésianisme hollandais*. Paris: Presses Universitaires de France, 1950.

#### WORKS BY REGIUS

*Fundamenta physices*. Amsterdam: Elsevier, 1646.

*Brevis explicatio mentis humanae sive animae rationalis ... a Notis Nobil. Cartesii vindicata*. Utrecht, Netherlands: N.p., 1648.

*Philosophia naturalis*. Amsterdam: Elsevier, 1654.

## WORKS ABOUT REGIUS

Rothschuh, Karl E. "Henricus Regius und Descartes: Neue Einblicke in die frühe Physiologie (1640–1641) des Regius." *Archives internationales d'Histoire des Sciences* 21 (82–83) (1968): 39–66.

Verbeek, Theo. *Descartes and the Dutch: Early Reactions to Cartesian Philosophy, 1637–1650*. Carbondale: Southern Illinois University Press, 1992.

Verbeek, Theo, ed. *Descartes et Regius: Autour de l'explication de l'esprit humain*. Amsterdam: Rodopi, 1993.

*Tad M. Schmaltz (2005)*

## REHMKE, JOHANNES

(1848–1930)

Johannes Rehmke, the German epistemologist, ontologist, and ethical philosopher, was born at Elmshorn in Schleswig-Holstein. He studied evangelical theology and philosophy at Kiel and Zürich from 1867 to 1871, receiving his doctorate in philosophy at Zürich in 1873. After some years as a high school teacher at St. Gallen, Rehmke was appointed unsalaried lecturer in philosophy at the University of Berlin in 1884. The following year he became professor of philosophy at the University of Greifswald, where he taught until 1921.

## THEORY OF KNOWLEDGE

Rehmke did not assume the existence of two worlds: a world, only indirectly knowable, of transsubjective objects, and an immediately knowable world, with intrasubjective perceptions and the like as contents. Rather, he asserted the existence of directly knowable real objects. This epistemological monism was a consequence of his ontological dualism of two essentially different kinds of being. Physical (material) beings are spatially extended and occupy a place; mental (immaterial) beings are not extended and have no place.

The nonspatial, placeless character of consciousness conflicts with the uncritical application to the subject of such concepts as "in" and "external," as exemplified in such terms as "intrasubjective" and "transsubjective"—in other words, "immanent" and "transcendent," or "content of consciousness" and "external object." Not only does consciousness not involve the having of a content; it does not involve any kind of having by means of a relation, in any event one that presupposes the existence of at least two realities separated from one another. On the contrary, knowing without any relation between diverse things is possible from the outset, as can be seen in self-consciousness. In self-consciousness only one thing is

given, the particular knowing consciousness as knowing itself and as being known by itself. Thus, Rehmke's proposition "Knowing is having without a relation" expresses the immediacy of all knowledge, including knowledge of the so-called external world, the world of objects outside the body.

In his *Logik oder Philosophie als Wissenslehre* (Leipzig, 1918), Rehmke sought to demonstrate the importance of the general or universal for the movement of knowledge toward clarity. In accord with his proof of the immediacy of cognition, he rejected as false the notion that thinking is an internal, that is, intramental, activity and even rejected the notion of thought activity because the purported activity never produces a change in objects. Thinking is not a "doing" but a "finding." If, for example, someone makes the judgment "A boiled crayfish is red," this observation signifies that he as thinker finds anew in the object the red known before. What is thus discovered in the object is never something single, an individual being, but something repeated, a universal.

Because the universal forms part of each particular object, it is something objective. If red is found in the crayfish, the logical subject of the judgment is not simply "(boiled) crayfish," but "red boiled crayfish." Consequently, every judgment, with respect to the universal discovered in the particular object, is logically analytic. Grammatically, with regard to the joining of the linguistic signs into a sentence, it is synthetic.

In its function as predicate of a judgment, an objective universal is called a concept. Every concept is thus a universal. Because of its objectivity, the universal as concept, despite its relation to the thinking subject, cannot be merely subjective. It is equally erroneous to confuse or to equate the concept, which is always bound up with a particular word, with that word, that is, with the phonic structure as linguistic sign.

The objectivity of the universal as a possible concept reveals the error in the phrase "concept formation." A concept (for example, "tree") is not first constructed by comparing several objects (for example, pines, beeches, and alders) by means of an "internal activity" of thought. The concept is presupposed in the very selection of objects of the same kind. Concept formation is really conceptual clarification, the determination of which characteristics in union constitute a concept already given. Clarity is the guiding notion in Rehmke's logic. He claimed that, in any deepening of knowledge, the universal as logical predicate helps consciousness to obtain clarity, and ultimately unquestionable clarity.

Rehmke's conception of logic, that is, philosophy as theory of knowledge, is linked with his notion of philosophy as fundamental science, expressed in his *Philosophie als Grundwissenschaft* (Frankfurt, 1910). Both theory of knowledge and fundamental science are genuine sciences, directed toward that which is simply given, that is, toward objects regardless of their being real or unreal. They are also in equal measure philosophy because they deal with the totality of the given, in contrast with the particular sciences, each of which deals with only a particular section of the world. Theory of knowledge deals with the given as that which is thought (known); fundamental science deals with it in regard to its most universal character. But while logic presupposes the concept "universal," and each special science presupposes its own fundamental concepts, the task of philosophy as fundamental science is to elucidate without prejudice precisely the basic "that which is most universal."

### THE TRADITIONAL ONTOLOGY

Theory of knowledge is not a fundamental science. Historically, it arose from an epistemological dualism, and as a consequence its form is faulty. In any case, it must presuppose the basic distinction between knower and other. Rehmke's painstaking ontological studies in *Philosophie als Grundwissenschaft* of the manifold "most universal" embrace five paired notions: (1) matter and consciousness, (2) the universal and the unique, (3) unity and simplicity, (4) the changeable and the unchangeable, (5) the real and the unreal. For Rehmke, of course, the first pair was primary. Beyond the merely negative description—immaterial, nonspatial, and place-less—the essence of the mental is completely determined by the concept of consciousness, or knowledge. Rehmke therefore opposed both materialism and idealism (spiritualism), as well as Spinozism.

Everything without exception proves to be either a unique thing (something that occurs only once, such as a unique tree) or a universal (something that is repeated, such as green or "treeness"). It follows that the unique and the universal do not exist without each other; indeed, objectively the universal belongs to the unique. Rehmke classified the unique into individuals (for example, individual trees) and units of individuals. He divided the latter into operational units (for example, an auto with a trailer) and living units (for example, a state). The universal is either a determination (such as angularity) or a relation (such as similarity). Rehmke attached great value to his recognition that many seemingly ontological con-

cepts, such as space, time, being, and value, are merely relational ones.

In connection with the third of his five pairs, unity and simplicity, Rehmke distinguished between individuals that are composed of individuals (and hence are ephemeral, passing) and individuals that are absolutely simple (and hence are everlasting). Examples of the latter are elementary particles and consciousness. Denying the theory of substance, he held that the individual is a union of its determinations (a body, for instance, is a union of size, shape, and location). He also analyzed each specific determination into determination as such (for instance, shape as such) and particularity.

Rehmke equated the fourth relationship, the changeable and the unchangeable, with the distinction between individual and universal. In this context he pointed out that the concept of change refers only to exchange of individual characteristics, with the determination as such (for instance, the shape as such) remaining the same.

Rehmke treated in detail the relationship between the real and the unreal. He defined the real as consisting in relationship of action. This enabled him to do justice to such properties of things as sweetness, which are often dismissed as merely subjective.

### PSYCHOLOGY AND ETHICS

In his *Lehrbuch der Allgemeinen Psychologie* (Frankfurt, 1894), Rehmke stressed that human consciousness (mind) is a simple, immaterial individual being, in a constant unity of action with an essentially different body. Thus, man is not a "double-being" individual. There are four general characteristics of consciousness: (1) determination of objects, each one directly perceived or imagined, even though the perception is mediated by the sense organs; (2) states (conditions), for example, delight or listlessness; (3) thought—either distinguishing (being aware of the distinct) or uniting (awareness of unity); (4) the subject, the determination of which establishes at the same time the unity of the ego. These determinations are not to be construed as mental activities.

Because of its intermittent character, volition, despite its relations with the above determinations, is not one of them. Rehmke's analysis of volition aided him in his solution of the problem of free will. He separated the problem into four parts, each of which is answerable: (1) Is an act of the will prevented or not? (2) Is the volition random or conditioned? (3) Is there a genuine possibility of choice, or is the will constrained? (4) Is the volition freely self-determined or not?

Rehmke's theory of the will constitutes the background for his ethics. He distinguished five forms of ethics—four false and one genuine. The ethics of shrewdness has to do with men “for themselves.” The ethics of the unity of control expresses duty as an “ought.” The ethics of the unity of life expresses duty as a “must” and comprises the ethics of society (in which unity as “being with one another” is a means to a selfish end) and the ethics of community (in which unity as “being for one another” is an end in itself). Finally, separating the merely social from the moral proper is the ethics of selfless love of one person as such “for another,” arising from his knowledge of himself as at one with the other.

**See also** Being; Consciousness; Determinism and Freedom; Epistemology; Ethics, History of; Idealism; Knowledge and Belief; Matter; Ontology, History of.

### Bibliography

#### ADDITIONAL WORKS BY REHMKE

- Unsere Gewissheit von der Aussenwelt* (Our certainty about the external world). Heilbronn, 1892.
- Grundlegung der Ethik als Wissenschaft* (Foundations of ethics as a science). Leipzig: Quelle and Meyer, 1925.
- Die Willensfreiheit* (The freedom of the will). Leipzig, 1925.
- Gesammelte philosophische Aufsätze* (Collected philosophical essays). Edited by Kurt Gassen. Erfurt: K. Stenger, 1928.
- “Selbstdarstellung.” In *Die Philosophie der Gegenwart in Selbstdarstellung*, edited by Raymund Schmidt, 7 vols., Vol. I, 177–200. Leipzig, 1921–1929.
- For a full bibliography see *Grundwissenschaft. Philosophische Zeitschrift der Johannes-Rehmke-Gesellschaft* 1 (1919): 72–88, and 10 (1931): 36–44.

#### WORKS ON REHMKE

- Heyde, J. E. *Grundwissenschaftliche Philosophie*. Leipzig, 1924.
- Heyde, J. E. “Johannes Rehmke.” *Grundwissenschaft. Philosophische Zeitschrift der Johannes-Rehmke-Gesellschaft* 10 (1931): 1–35.
- Heyde, J. E. “Johannes Rehmke.” *Zeitschrift für philosophische Forschung* (1947): 603–606.
- Heyde, J. E. “Johannes Rehmke 1848–1930.” *Philosophische Studien* 2 (1951): 260–271.
- “Johannes Rehmke.” In *Grundriss der Geschichte der Philosophie*, edited by F. Schneider, 319–329. Bonn, 1959.

*Joh's Erich Heyde* (1967)

Translated by Albert E. Blumberg

## REICH, WILHELM

(1897–1957)

Wilhelm Reich was an Austrian psychiatrist and social critic. After serving in the Austrian army during World

War I, Reich became a medical student. He obtained his M.D. from the University of Vienna in 1922 and worked for some time as assistant to Julius Wagner-Jauregg at the latter's psychiatric clinic. Even before his graduation Reich began practice as a psychoanalyst and soon came to occupy an influential position in the psychoanalytic movement. From 1924 to 1930 he conducted what came to be known as the Vienna Seminar for Psychoanalytic Therapy, the first organized attempt to devise a systematic and effective analytic technique.

Reich also founded and directed sex hygiene clinics among the industrial workers of Vienna and later, on a much larger scale, in Berlin and other German cities. During his years in Germany, Reich was a member of the Communist Party, and he attempted to integrate his work as a sex counselor within the broader revolutionary movement. Adolf Hitler's assumption of power forced Reich to flee to Denmark. His activities had always been viewed with suspicion by the leaders of the Communist Party, and Reich was finally expelled from the party after the publication of *Die Massenpsychologie des Faschismus* (Copenhagen, 1933), in which he repudiated the official communist theory about the nature of fascism and the factors leading to its victory in Germany. Also, by 1933 Reich's psychiatric views were so far removed from those of orthodox psychoanalysis that the Internationaler Psychoanalytischer Verlag handled and printed but did not “publish” (that is, refused its imprint to) the first edition of Reich's *Charakteranalyse*. The break with the psychoanalytic organization became official at the Lucerne conference of the International Psychoanalytic Association in 1934.

Attacks by orthodox psychiatrists made it necessary for Reich to leave Denmark for Sweden, but in Sweden too there was official hostility and suspicion. Reich therefore gladly accepted an invitation by the Norwegian psychologist and philosopher Harald Schjelderup to teach at the University of Oslo, where he also hoped to undertake various physiological experiments. Reich worked in Norway from 1934 to 1939. Among his students and patients at that time were the English educational reformer A. S. Neill, the American psychiatrist and pioneer in psychosomatic research T. B. Wolfe, and leading figures in Norwegian psychiatry, including Nic Hoel (Waal), Ola Raknes, and Odd Havrevold. The distinguished Norwegian novelist Sigurd Hoel was also closely associated with Reich at this time—in fact, he succeeded Reich as editor of the journal *Zeitschrift für politische Psychologie und Sexualökonomie*. In 1937 Reich became the victim of a campaign in sections of the Norwegian press. Although he

had a number of influential defenders and the government renewed his permit to stay in the country, he decided to move to New York City, where he resumed his psychiatric practice and trained numerous psychiatrists in the new technique that he had worked out during his stay in Scandinavia. Reich also lectured at the New School for Social Research from 1939 to 1941.

In the last years of his life Reich showed little interest in psychiatry, devoting all his energies to what he took to be his great discoveries in physics. In 1956 he was sentenced to two years' imprisonment for disobeying a government injunction. He died in Lewisburg Penitentiary in 1957. A brief account of the main events leading to Reich's imprisonment will be found in the last section of the present entry.

It will be convenient to distinguish three phases in Reich's career: (1) his work within the psychoanalytic movement, marked, however, by some significant departures from orthodox psychoanalysis—the rejection of symptom analysis in favor of what Reich called “character analysis,” the orgasm theory, and the attempt to understand the social function of sexual repression and neurosis; (2) Reich's efforts to relate neurotic attitudes to their somatic foundation and the development of what he called “character-analytic vegetotherapy”—a technique that constituted a drastic departure from all that preceded it; and (3) his theories about orgone energy—Reich's claim to have discovered a form of energy that is found in the atmosphere and also in the living organism and which can be concentrated in various ways, including the “orgone accumulator.” What Reich claimed during the third period is of no philosophical interest. If any of the assertions in question were true, they would be of great scientific interest; but, in fact, most professional physicists who have heard of the orgone theory have dismissed it as nonsense. In fairness to Reich it should be added that a really unbiased investigation of his physical theories remains to be undertaken.

We shall here be exclusively concerned with certain of the ideas advanced by Reich during the first two periods. Of interest to philosophers are Reich's views concerning the origin of religious and metaphysical needs, the relation between the individual and society and the possibility of social progress, and, above all, the implications of his psychiatry for certain aspects of the mind-body problem. It is regrettable that, partly because Reich's books and articles were not easily accessible and partly because the wild claims of his last years created widespread distrust of his entire work, the remarkable achievements of his second phase are relatively little-

known. To those who are put off by the recent metaphysical and pro-religious trends in psychiatry, as exhibited in the vogue of existentialist psychoanalysis and in the metapsychological speculations of Carl Jung and various Freudian analysts, Reich's concentration on the somatic basis of neurotic disturbances and the sexual problems and longings of human beings will come as a pleasant and refreshing change.

## THERAPEUTIC INNOVATIONS

The philosophically most interesting part of Reich's work is unquestionably what he called “the breakthrough into the vegetative realm,” that is, his attempt to determine the physiological basis of neurotic phenomena. However, first we should briefly describe Reich's earlier psychiatric work. In the early 1920s Freudian psychiatrists practiced what in retrospect came to be known as “symptom analysis.” Neurotic symptoms were regarded as foreign bodies in an otherwise psychologically healthy organism; they are expressions of a repressed infantile drive that has reappeared in a disguised form. The task of therapy is to eliminate the repression: The symptom is removed by bringing the repressed part of the personality into harmony with the rest of the ego. By his own account, Reich soon became dissatisfied with this approach. The traumatic experiences leading to repression and the repressed drives were to be elucidated by means of free association and dream interpretation, but in fact only very few patients were capable of giving their associations free rein. Furthermore, Reich was critical of the superficial criteria of “cure” current at that time. Patients were considered “cured” upon the disappearance or alleviation of the symptom of which they had complained. However, Reich believed that the elimination of symptoms is quite compatible with the continuation of a character disturbance. Also, he questioned the existence of “monosymptomatic neuroses”—neuroses with only one serious symptom. “There are no neurotic symptoms,” he later observed, “without a disturbance of the total character. Neurotic symptoms are, as it were, nothing but peaks of a mountain chain representing the neurotic character” (*The Function of the Orgasm*, p. 16). It was Reich's contention that, unless the characterological basis of a symptom has been eliminated, it or some equally troublesome symptom is likely to reappear.

On the few occasions on which either Reich or his associates at the Vienna Seminar appeared to achieve impressive and lasting improvements, this was invariably the result of the release of powerful dammed-up emotions like rage and hatred. Some years earlier, while work-

ing in Wagner-Jauregg's clinic, Reich had been struck by a catatonic who suddenly abandoned his stupor. "It was one great discharge of rage and aggression," Reich writes. "After the seizure had subsided he was clear and accessible. He assured me that his explosion had been a pleasurable experience, a state of happiness. He did not remember the previous stuporous phase.... It was very impressive, and could not be explained on the basis of the psychoanalytic theory of catatonia" (*The Function of the Orgasm*, pp. 43–44). Neurotics, too, showed noticeable improvement only when, instead of merely achieving an intellectual recognition of a repression, the impulse or emotion in question could actually be experienced. Such "liberations" were, however, infrequent and, what is more, they occurred more or less accidentally. An effective therapy would have to bring them about in a controlled fashion.

THE "CHARACTER ARMOR." Something should be said at this stage about Reich's concept of the "character armor" that came to play a central role in the technique of character analysis with which he gradually replaced the technique of symptom analysis. This concept was originally introduced in connection with certain cases of compulsion neurosis. Sigmund Freud had shown that compulsion symptoms always bind anxiety. If such a symptom is disturbed, the anxiety frequently appears. It does not, however, always appear—anxiety cannot usually be released in this way either in compulsion neuroses of long standing or in cases of chronic depression. Such patients appeared quite inaccessible. "Emotionally blocked compulsive characters gave associations in great numbers freely, but there never was a trace of affect. All therapeutic efforts bounced back, as it were, from a thick, hard wall" (*The Function of the Orgasm*, p. 114). These patients were "armored" against any attack. Over the years they had developed a set of attitudes whose function was to protect the individual against external injury (such as being hurt or rejected by other human beings) and to protect him against feeling his own repressed emotions, especially (though not exclusively) various kinds of destructiveness.

Reich introduced the term *character armor* to refer to the totality of the typical or chronic attitudes of this kind characterizing a given individual. It is, writes Reich, "as if the affective personality put on an armor, a rigid shell on which the knocks from the outer world as well as the inner demands rebound. This armor makes the individual less sensitive to unpleasure, but it also reduces his libidinal and aggressive motility and, with that, his capacity for pleasure and achievement" (*Character Analysis*, p.

310). Patients who do not suffer from a severe compulsion neurosis (and indeed most people growing up in a repressive environment) also have a character armor, but in their cases it can usually be attacked or broken down more easily.

The technique used to attack the character armor emphasizes the so-called negative transference. According to Reich, every patient has a deep mistrust of the treatment and feels strong hostility to the psychiatrist. Although patients wish to be cured, they also resent any attempt to disturb their "neurotic equilibrium." It is tempting for the analyst to shy away from these negative reactions, since it takes a great deal of strength and composure to bear the often furious hatred that is released when the armor begins to "crack." Nevertheless, it is precisely this negative reaction that can and must be used as the foundation of the treatment. The patient must feel free to criticize the analyst, and any attitudes that mask his hostility have to be broken down. Reference to the case of a "passive-feminine young man with hysterical symptoms" may give some idea of what this technique is like. The patient was excessively polite and, because of his fears, extremely sly. He always yielded and produced abundant material, but without any inner conviction. "Instead of discussing this material," Reich reports,

I only kept pointing out his politeness as a defense against me and any really affective insight. As time went on, his hidden aggression appeared increasingly in his dreams. As the politeness decreased, he became offensive. In other words, *the politeness had been warding off the hatred*. I let the hatred come out fully by destroying every defense mechanism against it. The hatred up to that time had been unconscious. Hatred and politeness were antitheses, and at the same time the over-politeness was a disguised manifestation of hatred. (*The Function of the Orgasm*, p. 117)

If in this way repressed emotions are released and the patient actually experiences them, it is unnecessary to persuade him that he "really," "unconsciously" feels this or that. "The patient no longer talked about his hatred, he felt it; he could not escape it as long as his armor was being correctly taken apart" (p. 146).

The armor, according to Reich, varies from patient to patient, depending on his individual history, and the technique of destroying it has to be fitted to the individual case. The armor may be viewed as consisting of several layers. These layers, in Reich's words, "may be compared to geological or archaeological strata which,

similarly, are solidified history. A conflict which has been active at a certain period of life always leaves its trace in the character, in the form of a rigidity” (pp. 121–122). The neurosis of each patient has a specific structure that corresponds to its historical development, but in reverse order: “that which had been repressed latest in childhood was found to lie nearest the surface” (p. 121).

Anger and hate are not the only emotions bound by the character armor. Although destructiveness has to be emphasized and liberated in the early stages of the treatment, eventually genuine love and tenderness that had to be suppressed will also be released. The destructiveness, in the last resort, is “nothing but anger about frustration in general and denial of sexual gratification in particular” (p. 124). Destructive tendencies are most frequently “reactions to disappointment in love or to loss of love.” An organism that has been freed of its dammed-up destructiveness becomes once again capable of love. Reich referred to persons who are unarmored and who possess the capacity for love in the fullest sense as “genital characters”; and the goal of therapy is to change the patient’s neurotic character into a genital structure. According to Reich, the “energy” that nourishes neurotic symptoms and various destructive attitudes can be adequately discharged only in fully satisfactory sexual intercourse. A person with a genital character, unlike the neurotic, possesses “orgastic potency.” This Reich defined as “the capacity for surrender to the flow of energy in the orgasm without any inhibitions; the capacity for complete discharge of all dammed-up sexual excitation through involuntary pleasurable contractions ... free of anxiety and unpleasure and unaccompanied by phantasies” (*The Function of the Orgasm*, p. 79).

An individual with a genital character has undisturbed contact with his own drives and with his environment and, as a consequence, he has no need for any of the endless variety of substitute contacts and substitute gratifications of the neurotic individual. He, too, may not succeed in achieving a happy existence, since this depends on a great many factors, not all of which are within his control, but he will at least not be hampered in his struggle for happiness by irrational and destructive emotions or by excessive respect for the institutions of a life-denying society.

Reich vigorously repudiated the suggestion that, either in his therapy or in his social philosophy, his goal was a world “containing nothing but pleasure.” The function of the armor, he observed, is to protect against pain, and in breaking it down, Reich’s therapy aimed at reestablishing the capacity to feel pain as well as pleasure. “Plea-

sure and joie de vivre are inconceivable without fight, without painful experiences and without unpleasurable struggling with oneself” (p. 173). The goal is not a positive “hedonic balance” that, for all one can prove to the contrary, might be more effectively achieved by a life of monasticism but “full vitality in all possible situations of life.” The capacity to take happiness and to give love goes hand in hand with “the capacity of tolerating unpleasure and pain without fleeing disillusioned into a state of rigidity.”

**REPRESSIONS AND CHRONIC MUSCULAR RIGIDITIES.** Reich was led to his study of what he calls the “physiological anchoring” of neurotic conflicts and traumatic experiences partly as a result of his fundamentally materialistic orientation and partly because of the special attention paid in his technique of character analysis to the *manner* in which patients talked and acted. It is a mistake, he said, to regard rage and love (or any other emotion) as events “in the mind.” They *are* physiological processes, and if an emotion is repressed, there must be some physiological mechanism by whose means the energy in question is “bound.” Furthermore, Reich was convinced that if an adult’s neurotic character attitude is the result of childhood experiences, this can be so only if the person’s organism has in some way been chronically altered. The employment of “theoretical” terms such as “Id” and “unconscious” can easily lead to pseudoexplanations in this context. To say, for example, that a repressed childhood conflict exerts its influence “from the unconscious” may call attention to a suspected causal relation between the childhood experience and the present difficulties of the individual, but beyond that it simply amounts to admitting that one does not know how the influence in question is exerted. On occasions, it is true, Freud himself said as much and expressed his hope that some day explanations in terms of unconscious conflicts would be given a physiological meaning. At other times, however, Freud treated his theoretical terms as if they designated real and eternally inaccessible entities; and many of Freud’s followers, according to Reich, became metaphysicians whose theorizing was euphemistically labeled “metapsychology.”

Perhaps of greater influence than these general reflections was Reich’s interest in the “how” of the patient’s communications. The infantile structure, Reich observes in one place, is “conserved” in what an individual does as well as in the way in which he acts, talks, and thinks (*Character Analysis*, p. 188). Elsewhere Reich explains that he made himself independent of the so-called fundamental psychoanalytic rule (“to say everything that comes to

mind”), since it was impracticable with most patients, and that instead he took as “point of attack not only what the patient said, but everything he presented, particularly the manner of his communications or of his silence. Patients who kept silent were also communicating, were expressing something that gradually could be understood and handled” (*The Function of the Orgasm*, p. 145). It became increasingly evident to him, Reich adds, that “the form of behavior and communications, was much more essential than what the patient related. Words can lie. *The mode of expression never lies.*”

Special attention to the “how” of a patient’s behavior very naturally led to close observation of the changes in his organism during and after the release of repressed emotions. Reich’s earlier clinical reports already contained remarks about the awkwardness and the rigid movements of certain types of patients. However, it was not until the early 1930s that he began to elucidate the precise role played by muscular rigidities in the binding of impulses and emotions that had to be suppressed. The following extracts describing the beginning of a treatment in 1933 will perhaps convey better than any definition what Reich meant by the “physiological anchoring” of affects:

In Copenhagen in 1933, I treated a man who put up especially strong resistances against the uncovering of his passive-homosexual phantasies. This resistance was manifested in an extreme attitude of stiffness of the neck. ... After an energetic attack upon his resistance he suddenly gave in, but in a rather alarming manner. For three days, he presented severe manifestations of vegetative shock. The color of his face kept changing rapidly from white to yellow or blue; the skin was mottled and of various tints; he had severe pains in the neck and the occiput; the heartbeat was rapid; he had diarrhea, felt worn out, and seemed to have lost hold. ... *Affects had broken through somatically after the patient had yielded in a psychic defense attitude.* The stiff neck, expressing an attitude of tense masculinity, apparently had bound vegetative energies which now broke loose in an uncontrolled and disordered fashion. ... It was the musculature that served this inhibitory function. When the muscles of the neck relaxed, powerful impulses broke through, as if propelled by a spring. (*The Function of the Orgasm*, pp. 239–240)

This and other cases led Reich to a systematic study of chronic muscular rigidities and their relation to neurotic character attitudes. He reached the conclusion that “every neurotic is muscularly dystonic and every cure is directly reflected in a change of muscular habitus” (*Character Analysis*, pp. 311–312). Chronic muscular rigidities or spasms are found all over the bodies of the patients: in the forehead, around the mouth and in the chin, in the throat, the shoulders, the chest, the abdomen, the pelvis and thighs, and many other places. The rigid expression in the eyes of many patients, their chronic “stare,” is the result of a chronic rigidity in the lid muscles. The breathing of neurotic individuals is disturbed in comparison with the natural and free respiration of emotionally healthy people. Reich referred to the totality of these chronic muscular rigidities that an individual develops as the “muscular armor.”

Reich emphasized that it is muscle groups rather than individual muscles that become spastic—muscle groups which jointly serve a certain function, for example, to suppress the impulse to cry. Not only do the lower lips become tense in this event but also “the whole musculature of the mouth, the jaw and the throat; that is, all the muscles which, as a functional unit, become active in the process of crying” (*The Function of the Orgasm*, p. 269). In discussing the spasms frequently found in the mouth, chin, and neck, Reich enlarges on the tensions set up by the stifling of impulses to cry:

Many people have a mask-like facial expression. The chin is pushed forward and looks broad; the neck below the chin is “lifeless.” The lateral neck muscles which go to the breastbone stand out as thick cords; the muscles under the chin are tense. Such patients often suffer from nausea. Their voice is usually low, monotonous, “thin.” (*The Function of the Orgasm*, p. 271)

This is not the place to discuss in detail other of the typical rigidities that make up the muscular armor. The interested reader will find these described in various of the publications devoted to the new technique.

Upon discovering the muscular spasms and their relation to suppressed impulses and emotions, Reich devised various ways of attacking or “dissolving” them directly. In working on tensions in and around the eyes, for example, it is frequently possible to release a great deal of anxiety; in loosening up and encouraging the movement of certain muscles around the mouth, suppressed feelings of disgust can be liberated; by suitable work on the chin, it is possible, in Reich’s words, “to set free an unbelievable amount of anger.” Reich writes that he had



previously been able to bring about the release of repressed impulses and emotions by way of dissolving purely characterological inhibitions and attitudes. Now, however, “the break-through of biological energy was more complete, more forceful, more thoroughly experienced, and it occurred more *rapidly*. Also, it was accompanied in many patients by a spontaneous dissolution of the characterological inhibitions” (p. 241). Reich warns, however, that it is not possible to dispense with work on character attitudes. “Everyday practise soon teaches one,” he writes, “that it is not permissible to exclude one form of work at the expense of the other” (p. 293). With some patients work on the muscular rigidities will predominate from the beginning; with others, work on the character attitudes; but in all cases work on the muscular armor becomes more important in the later stages of the treatment.

**MIND-BODY PROBLEM.** The facts he discovered about chronic muscular rigidities and their relation to character attitudes and repressed emotions, Reich maintained, required the abandonment of the dualistic theories about body and mind tacitly or explicitly accepted by many psychologists and most psychoanalysts. It is a mistake to regard the muscular rigidity as a mere accompaniment or as an effect of the corresponding character attitude: It is “its somatic side and the basis for its continued existence” (*The Function of the Orgasm*, p. 269). The rigidity of a muscle group and the corresponding attitude serve the same function, namely, that of holding back a repressed emotion. The muscular armor and the character armor may therefore be said to be “functionally” identical. The only tenable answer to the body-mind problem, according to Reich (who quotes Julien Offray de La Mettrie as anticipating his position), is a materialistic form of the identity theory.

Reich’s identity theory is materialistic, not in the sense that introspection is regarded as illusory or as devoid of scientific value but in holding that a change in a person’s character, or indeed any change in a human being, cannot come about without appropriate physiological changes. The notion, writes Reich, that “the psychic apparatus functions by itself and influences the somatic apparatus is not in keeping with the facts” (*ibid.*, p. 313). Even an idea such as that of going to sleep will not “exert a somatic influence unless it is already the expression of a vegetative impulse” (*ibid.*). This conclusion, Reich insists, is not contradicted by the observation that a patient (or anybody) feels relieved when a previously repressed idea or impulse is allowed to become conscious. “We used to say,” writes Reich, “that it is a matter of a dis-

charge of psychic energy which previously was bound” (*Character Analysis*, p. 311).

Closer examination will show in such a case that both the tension and the relaxation are clearly observable somatic processes. What is introspectively felt as tension and as relief are in fact certain fairly typical rigidities and relaxations of muscles—in the forehead, in the eyes, and elsewhere in the body. Both Reich and his translator, T. P. Wolfe, insist that the issue between dualism and the identity theory is not merely a question of alternative languages but makes a difference to therapeutic practice and further research. Wolfe in particular claims that only a theory of “psychosomatic identity” makes sense of the vast array of facts that had accumulated in psychosomatic studies by 1940 and that only such a theory can provide a fruitful method of research (*The Function of the Orgasm*, pp. x and xiii).

There are two very different questions that may be raised about all of this. One may ask whether, granting that Reich has hit upon something interesting and important in connection with muscular rigidities, their origin, and their possible dissolution, an identity theory is the only philosophical position that can accommodate these facts. More fundamentally, one may raise the question of whether Reich’s empirical claims about the muscular armor are true in the first place.

As to the first of these questions, it should be pointed out that when Reich speaks of the “*functional* identity” of the character and the muscular armor, he does not seem to mean by “identity” anything as strong as has been claimed by philosophical defenders of the identity theory. To say that a certain character attitude and a certain muscular rigidity have the same function, for example, that of binding anxiety or anger, is not anything that a dualist is required to deny. It is certainly compatible with, but it does not by itself imply, the claim that the character attitude and the muscular rigidity are two aspects of the same phenomenon. It might be argued that Reich’s work on the connection between muscular rigidities and character attitudes, rather than proving any traditional version of the identity theory, shows the inadequacy of interactionistic forms of dualism. Interactionism, in allowing only for causal relations between physical and psychological phenomena, could not do justice to the intimate relations between muscular rigidities and character attitudes to which Reich has called attention. There is no reason to suppose, however, that a more open-minded form of dualism, which would not restrict the relations between body and mind to one simple type, could not accommodate the facts in question.

In the present entry we cannot attempt to answer the second of our two questions—whether Reich’s empirical claims about the muscular armor can in fact be sustained. Perhaps, however, it is permissible to remark, especially since this part of Reich’s work has received so much less publicity than his orgone theory, that psychiatrists and others who have some firsthand knowledge of it have generally been enthusiastic. This includes persons who have observed and treated children in the light of Reich’s account of the muscular armor. Since the process of repression as well as the process of cure would, on almost any theory, be most readily observable in children, confirmations (or disconfirmations) here would seem to be of special significance.

## CULTURE, SOCIETY, AND CHARACTER STRUCTURE

**CULTURE, MORALITY, AND THE DEATH INSTINCT.** On the basis of both his clinical observations and his very extensive social work, Reich maintained that there is nothing more deadly than to be subjected to the moralistic and authoritarian upbringing which is or which was until very recently the lot of the great majority of children all over the world. The preaching and the antisexual moralism of the religious home and the authoritarian character of the conventional school stifle every vital impulse in the child. Insofar as traditional education is successful, it produces human beings with a craving for authority, a fear of responsibility, mystical longings, impotent rebelliousness, and pathological drives of all kinds. The “morals” fostered by religious mysticism and slavishly followed by many who no longer believe in religion “create the very perverted sexual life which it presumes to regulate moralistically; and the elimination of these ‘morals’ is the prerequisite for an elimination of that immorality which it tries in vain to fight” (*The Mass Psychology of Fascism*, p. 156).

There is nevertheless an important element of truth in the contention of conservative ideology that if one were to “eliminate morals,” the “animal instincts” would gain the upper hand, and that this would lead to social chaos. What is true in this contention is that the average person in our culture carries within himself an “unconscious inferno,” and while his perverse and destructive impulses are not in most cases adequately controlled by moral inhibitions, they would presumably dominate personal and social life to an even greater extent in the absence of moral regulations. This fact makes it clear that any transition from an authoritarian to a rational self-governing society must be gradual and cannot be accom-

plished by simply telling people, *as they now are*, to live according to their impulses. It does not, however, provide a justification for an ascetic morality or for the usual conservative theory that maintains that culture is based on sexual repression.

The conservative theorist errs in assuming that the antisocial impulses are “absolute and biologically given” (*The Sexual Revolution*, p. 20). This view is advocated not only in the writings of religious moralists and others to whom Reich contemptuously referred as “uplifters” or “guardians of the higher values” but also in many of the later writings of Freud and those of Freud’s followers who accepted the theory of the death instinct. Accordingly, Reich devoted much effort to a very detailed attack on the theory of the death instinct, especially as it is applied to human society and culture in Freud’s *Civilization and Its Discontents*.

On Freud’s view, Thanatos, or the striving for peace and extinction, is just as much biologically given as Eros, or the sexual strivings. Although the death instinct itself cannot be perceived, it manifests itself in a great many ways—in various forms of aggression, in self-destructiveness, and in the masochistic “need for punishment.” It also accounts for the resistances put up by patients against getting well. According to Reich, however, both clinical experience and observation of children show that the phenomena which supposedly prove the death instinct are “secondary formations,” the products of the neurosis, and not “primary” and “biological” like the sexual instinct or the need for food. Investigation reveals that suicide is either an unconscious revenge upon another person or a way of escaping the pressure of a situation that has become overwhelming. The neurotic fear of and concern with death that is frequently found in quite young people can in every case be reduced to a fear of catastrophe, and this, in turn, to genital anxiety. As for aggressiveness, Reich claimed that the proponents of the death instinct did not sufficiently distinguish between perfectly healthy forms and those which are sadistic and destructive. The former are intimately connected with life-affirming tendencies, and the latter are always reactions of the organism to the denial of the gratification of a vital need.

Reich equally denied that there is any evidence whatsoever for the theory of “primary masochism.” All clinical observations support Freud’s earlier theory that patients “had come to grief as a result of their *fear* of punishment for sexual behavior and not as a result of any *desire* to be punished for it” (*The Function of the Orgasm*, pp. 103–104). The theory of the death instinct, furthermore,

is therapeutically sterile and offers an excellent excuse for one's inability to handle a difficult resistance. In addition to providing an alibi for therapeutic failures, it serves the same function as the discredited biologicistic theory of congenital criminality or the view of Magnus Hirschfeld that exhibitionism is due to special "exhibitionistic hormones": All such views shift problems from the social to the biological realm, where nothing can—and need—be done about them.

Conservative theorists who maintain that there is an antithesis between sexuality and work fail to distinguish between "compulsive-unpleasurable" work, which is indeed regarded as a burdensome duty, and "natural joyful work," which frequently requires discipline but which is nevertheless a pleasurable gratification of a need. Reich regards as especially significant his observations on patients who achieved sexual happiness. He reports that those who, because of neurotic disturbances, had not been working, began to feel a strong need for some vital work. Those who had been engaged in work that was intrinsically interesting now blossomed and gave full rein to their talents. In some cases, however, there was a complete breakdown of work. This at first seemed to confirm the view of the antisexual moralists, but closer inspection showed that these people had previously been driven by a compulsive sense of duty and that what they rebelled against was empty and mechanical work, and not work as such. Their aversion was to pleasureless work, and their impulses were by no means antisocial. Just as society rewards some highly antisocial activities with fame and honor, Reich remarks, so "there are highly valuable, even culturally important traits and impulses which have to be repressed for considerations of material survival" (*The Function of the Orgasm*, p. 150). If there were more human beings with a genital character, this would not result in the end of "civilization," but it would in all probability lead to radical changes in the ways in which the world's work is done.

Reich concluded that civilization and culture do not depend on instinctual repression. If authoritarian education were abolished and if children grew up in a sex-affirmative environment, people would be more, and not less, peaceful and cooperative. Some types of work, namely, those in which only a person with a compulsive character can take any interest, would indeed suffer, but the arts and sciences would in all likelihood flourish as never before. Reich was not an irrationalist in any sense of the word, and like Freud he favored "the primacy of the intellect," adding, however, that the full utilization of a person's intellectual capacities presupposes "an orderly libido

economy." "Genital and intellectual primacy have the same mutual relationship as have sexual stasis and neurosis, guilt feelings and religion, hysteria and superstition" (*Character Analysis*, p. 170).

**SOCIETY AND CHARACTER STRUCTURE.** Freudian social theory, insofar as it existed at all when Reich began his elaborate critique of what he called "authoritarian" society, was vitiated by its "biologism" as well as its "psychological atomism," or, as Reich also called it, a "feudal individualistic psychology." By "biologism" Reich meant the tendency to treat as universal and biologically inevitable attitudes and impulses that were determined by cultural conditions. When he spoke of Freud's "psychological atomism," Reich referred to the tendency to treat individual patients and their families in isolation from the social environment that had in fact a great deal to do with their tribulations.

Rejecting Freud's biologism and accepting the early Freudian view that neurosis is basically the result of the conflict between instinctual needs and the reality which frustrates them, Reich naturally asked whether and how this frustrating reality could be significantly altered. His work at the sex hygiene clinics, furthermore, had convinced him that neuroses were by no means the fads of middle-class women who did not know what to do with their time but were emotionally crippling illnesses of almost epidemic proportions. Contrary to the assertions of the more doctrinaire and narrow-minded Marxists, there could be no doubt in Reich's view that "sexual repression, biological rigidity, moralism and puritanism are ubiquitous" and not confined to certain classes or groups of the population (*The Function of the Orgasm*, p. xxiii). The vast majority of people suffering from psychological disturbances cannot, however, be reached by individual therapy, disregarding here all the difficulties and limitations of such therapy when it is available.

If one is to do anything about this deplorable state of affairs, one must first achieve an understanding of the precise relations between society and the individual and, more specifically, between social institutions and neurotic disturbances. "Society," Reich writes, "is not the result of a certain psychic structure, but the reverse is true: character structure is the result of a certain society" ("Character and Society," p. 254). The ideology of a given society can anchor itself only in a certain character structure, and the institutions of the society serve the function of producing this character structure. If, as in all authoritarian societies, a minority holds economic and political power, it also has the power to form ideology and structure. As a

consequence, in authoritarian society, the thinking and the structure of the majority of people “corresponds to the interests of the political and economic rulers” (*The Sexual Revolution*, p. xx). The majority of human beings (Reich is writing in 1936) are “suppressed and exploited and spend most of their working hours doing monotonous and mechanical labor which they cannot help regarding as a loathsome duty.” How is it possible that “people can bear it, that they are unable to change it, that they seem to endure in silence the suffering it imposes on them?” (“Character and Society,” p. 252). They can bear their fate because the ruling economic system is “anchored in the psychic structure of the very people who are suppressed” (*People in Trouble*, p. 100).

The most important structure-forming institutions in authoritarian society are the authoritarian family, the authoritarian school, and religion. “From infancy on,” writes Reich, “people are trained to be falsely modest, self-effacing and mechanically obedient, trained to suppress their natural instinctual energies” (“Character and Society,” p. 252). In this way children become subservient to their parents and people in general “subservient to the authoritarian state power and capitalistic exploitation” (*People in Trouble*, p. 99). The most powerful instrument in achieving this mass structure is sexual repression, which is fostered in the home, in the school, and above all through the influence of religious moralism. The major mechanisms of sexual suppression in Christian countries are the prohibition of infantile masturbation, the prevention of sexual gratification in adolescence, and the institution of compulsorily lifelong monogamy, accompanied by the belief that the function of sexuality is procreation rather than pleasure. The parents who punish children for masturbating and who do their best to prevent adolescents from having a full sex life are unwittingly carrying out the purpose of the ruling powers.

There is something plausible about Reich’s contention that an atomistic psychology, no matter how correctly it may determine the causes of mental health and illness, will not by itself explain why various institutions that are plainly inimical to life and happiness nevertheless flourish and receive the support of all the major official and unofficial agencies of society. However, it is not entirely clear what he means by his claim that character structure is the result of social structure and, more specifically, that the “function” of sex-denying institutions is to make the masses helpless and dependent. Although he occasionally uses the word *purpose*, Reich is presumably asserting the existence of a “latent” rather than a “manifest” function, to use the terminology introduced by R. K.

Merton. While it may be plausibly argued that some rulers, like Joseph Stalin and certain church figures, have been aware of the connection between sexual suppression and such “desirable” traits as obedience and uncritical acceptance of the status quo, it would be farfetched to hold that either in capitalistic or in other societies the ruling circles deliberately support sex-denying institutions in order to perpetuate their power and privileges. But if the rulers are not conscious of the causal connection between sexual suppression and the submissive traits it produces, in what sense is a reference to their interest an explanation of the institutions in question? It is tempting to speak here of an “unconscious knowledge” or “unconscious realization” that sexual suppression produces submissiveness, but it is far from clear what these expressions would mean.

Reich’s views about the relation between the ideology that prevails in a society and the interests of the holders of power has obvious affinities with Marxism, and in fact a number of Marxist writers of the late 1920s and early 1930s hailed his account of the social function of sexual repression as a valuable supplement to historical materialism. However, the most influential Marxist ideologists, socialist as well as communist, rejected Reich’s account and also strongly opposed his work in his sex hygiene clinics. In his turn, Reich repudiated what he called the “economism” of Marxist theory as emphatically as he attacked the atomism of psychoanalysis. “Marxists again and again argued,” he recalls, “that the sexual etiology of the neuroses was a bourgeois fancy idea, that only ‘material want’ caused neuroses ... as if the sexual want were not a ‘material’ one: It was not the ‘material want’ in the sense of the Marxian theorists that caused the neuroses, but the neuroses of these people robbed them of their ability to do anything sensible about their needs, actually to do something constructive about their situation, to stand the competition on the labor market, to get together with others in similar social circumstances, to keep a cool head to think things out.” (*The Function of the Orgasm*, pp. 56–57).

Moreover, just as it is wrong to think that neuroses are (except very indirectly) caused by economic hardships, so it is a mistake to suppose that the social and political actions of the working classes can be predicted on the basis of their economic interests alone. Factors such as mystical and sexual longings and perverse sadistic fantasies may exert very powerful influences, as Hitler, unlike the communist, socialist, and liberal politicians, understood only too well. Fascism, to take but one example, is very incompletely characterized as a movement

engineered by capitalists to prevent the establishment of socialism. At least the German variety of fascism differed from other reactionary movements in that it was “supported and championed by masses of people” (*The Mass Psychology of Fascism*, p. ix). Marxist theory, which assumes that with few exceptions the underprivileged will be guided by their rational economic interests, is incapable of accounting for such a phenomenon.

## THE STRUGGLE AGAINST RELIGION

**MYSTICAL FEELINGS AND SEXUAL INHIBITIONS.** According to Reich, both Karl Marx and Freud made significant contributions to our understanding of religion. Patriarchal religions are always politically reactionary, and Marxists are perfectly right in pointing out that “in every class society they are in the service of the powers that be” (p. 124). Freud, too, was correct in his view that the idea of God derives from the idea of the father and, more generally, that “the psychic contents of religion stem from the infantile family situation” (*ibid.*). Granting all of this, there remains a question that is not answered by the Marxist or the Freudian account, or by any of the great eighteenth-century critics of religion. Indeed, it is a question that most of these writers did not even raise but which must be asked and answered if one is to have an adequate comprehension of religion. How are we to account for the fact that “religious ideas are invested with such intense feelings”? What explains the “enormous emotional power of mysticism” (p. 122)? Or, using Reich’s favorite terminology, what is the “energy” that enables religions to gain such a firm hold on people? What is it that compels human beings not only to accept the idea of a pleasure-prohibiting, all-seeing God and the ideologies of sin and punishment, and “not to feel them as a burden but, on the contrary, to uphold and fervently defend them, at the sacrifice of their most primitive life interests?” (*The Mass Psychology of Facism*, p. 124).

Reich is strongly opposed to the tendency of “emancipated” unbelievers to dismiss religions as nothing more than the fancies of silly and ignorant people. He insists that a study of religious people—of the content of their emotions and beliefs, of the ways in which these are implanted, and of the function that they fulfill in their psychological economy—is highly rewarding. It sheds light on many other phenomena, including, for example, the psychological basis of fascism and of reactionary political movements. Such a study also explains why, by and large, free-thought propaganda is so unsuccessful in spite of the fact that from a purely rational point of view the positions defended by freethinkers are vastly superior

to the religious claims—something that is not altogether unknown among believers. Above all, a happy life for the majority of humankind is impossible unless the power of religion is broken, unless one can prevent “the mystical infestation of the masses” (p. 161).

However, in order to be effective in “the relentless fight against mysticism,” one must have a full comprehension of its origin and its psychological sources of strength so that one can meet its “artful apparatus . . . with adequate counter-measures” (p. 152). To suppose that mystical attitudes become anchored in human beings simply as a result of intellectual indoctrination is a naive and dangerous mistake. It should be noted that Reich sharply distinguishes mysticism from primitive animism. The latter is best regarded as bad science. Reich does not offer an explicit definition of “mysticism,” but it seems clear from his various writings on the subject that mysticism in the “strict and wider sense” is characterized by the belief (or feeling) that the ordinary world of physical objects and human emotions is *not enough* and the related view that there are some grand truths which human beings can come to know by nonscientific or superscientific means. Various nontheological systems of metaphysics and ontology, as well as the standpoint of those who deny that psychology can properly be a natural science (Reich is specially scathing in his comments about Ludwig Klages and Karl Jaspers), are treated by him as forms of mysticism.

The most basic feature of what Reich variously calls “religious excitations” or “mystical feelings” is that they are “at one and the same time *anti*-sexual and a *substitute* for sexuality” (p. 125). Reich claims that this conclusion is borne out by the close observation of genuinely religious people (as contrasted with those who merely pretend belief for purposes of personal gain and advancement); by character-analytic treatment of religious individuals and patients having mystical feelings of any kind; by observation of children, especially those suffering from prayer compulsions; by the writings of the mystics themselves; and also by what is known about the changes that occurred when social organization passed from matriarchy to patriarchy and class society.

Biologically, the religious individual is subject to states of sexual tension like any other living being. However, as a consequence of his sex-negating upbringing and especially his fear of punishment, he has lost the capacity for normal sexual stimulation and gratification. The result of this is that he suffers from a chronic state of excessive somatic excitation. The more thorough his religious education has been, the more it appears to him that

happiness is not attainable for him in this world and, in the long run, it does not even seem desirable any more. However, he remains a biological organism and hence cannot completely renounce the goals of “happiness, relaxation and satisfaction.” In these circumstances all he can do is seek “the *illusory* happiness provided by the religious *forepleasure* excitations” (*The Mass Psychology of Fascism*, p. 126).

The “somatic suffering” of the religious person creates in him the need for consolation and help from outside himself, particularly in his fight against what he terms the “evil instincts,” which in turn are identified with the “evils of the flesh.” His religious ideas enable him to attain a state of “vegetative excitation which resembles gratification but does not, in reality, bring about somatic relaxation” (p. 127). Not even religious ecstasies bring about anything comparable to the orgasmic relief of a satisfying sexual experience. What the religious person calls his longing for “delivery from sin” is in fact a longing for relief from sexual tension. To people who cannot achieve sexual gratification, sexual excitation gradually and inevitably becomes something “torturing and destructive.” In this way the religious conception of sex as evil and debasing has its foundation in real somatic processes. People who feel a disgust for their body quite naturally develop obsessive concepts of “purity” and “perfection” (p. 144).

It would lead too far afield to discuss here the various ways in which, according to Reich, the “mystical idea of God” becomes anchored in people. These mechanisms may vary in detail, but they all involve the implanting of sexual anxieties; and Reich concludes that from the point of view of energy, mystical feelings are “sexual excitations which have changed their content and goal.” The energy of these emotions is the energy of natural sexuality that has become transformed and attached to mystical, psychic contents. Religious patients, upon establishing a fully satisfying sex life, invariably lose their God-fixation.

Once one comprehends the nature of “religious excitations,” it becomes clear why the free-thought movement “cannot make itself felt as a counter-force” (p. 147). Aside from the fact that in many countries the churches enjoy the support of the state and that generally the mass information media are grossly biased in favor of religion and religious morality, the impact of free-thought propaganda is limited because it relies almost exclusively on intellectual arguments. These are not, indeed, a negligible factor, but they are no match for the “most powerful emotion” on which the mass-psychological influence of religious institutions is based: sexual anxiety and sexual

repression. People with a religious upbringing who, as a result of the study of science and philosophy, have turned into unbelievers very frequently retain religious longings and emotions. Some of them even continue to pray compulsively. This does not prove, as some advocates of religion argue, that religious needs are “eternal and ineradicable.” It does, however, show that “while the religious feeling is opposed by the power of the intellect, its sources have not been touched” (p. 152).

The fight against religion is nevertheless far from hopeless. Mysticism can be eradicated if, in addition to depriving the churches of their “evil right of preparing the children’s minds for the reception of reactionary ideologies” (p. 148), one is guided in the struggle by one’s knowledge that mysticism stems from inhibited sexuality. From this insight it follows incontrovertibly that “full sexual consciousness and a natural regulation of sexual life mean the end of mystical feelings of any kind, that, in other words, natural sexuality is the deadly enemy of mystical religion” (p. 152). Any social efforts that are directed toward making people affirm their sexual rights will ipso facto weaken the forces of mysticism. The most good can be done with children and adolescents. Reich gives numerous instances from his experience in Germany of the “burning interest” of children in sexual questions that made even the most enlightened adults ashamed of their prudishness and hesitation. “Once children and adolescents are reached on a mass basis through their sexual interests,” there will be a “powerful counterweight against the reactionary forces” (p. 169). As for those people who are too old to have their structure basically altered, it is still all to the good to bring “silent suffering to the surface.” They might then be less likely to become instruments in the process of maiming their own children, and they will not continue to support sex-repressive laws.

**THE GREAT CULTURAL REVOLUTION.** Reich never abandoned the conviction he had reached during his Marxist phase that individual therapy is socially insignificant and that “alteration of the social structure is a prerequisite for an alteration of the psychic structure on a mass scale” (“Character and Society,” p. 255). However, after his separation from organized Marxism, he gradually came to the conclusion that political action was of little consequence and that it was a grave error to judge social developments primarily in terms of a rigid, clear-cut class war. If one is not blinded by the political slogans of an earlier age, one cannot help noticing that we are in the midst of a “deep-reaching revolution of cultural living” (*The Sexual Revolution*, p. xiv).

It is a revolution “without parades, uniforms, drum or cannon salutes,” but, unlike the Russian Revolution of 1917, which was merely “politico-ideological,” it is a “genuine social revolution” (*The Mass Psychology of Fascism*, p. 201). It is not a revolution by the proletariat against the bourgeoisie, and it remains to be seen what major economic changes will accompany it. What is happening is that “the senses of the animal, man, for his natural life functions are awakening from a sleep of thousands of years” (p. xiv). Ever since the beginning of the century, numerous social factors have been operating in the direction of freedom and health. These factors include the creation of huge industrial plants with vast armies of workers of both sexes and the gradual undermining of the authoritarian parental home. There has been a “thorough disintegration of the moralistic ascetic forms of living,” and this “objective loosening of the reactionary fetters on sexuality cannot be undone” (p. 164), regardless of how vociferously the churches and their conscious or unconscious allies continue to preach the old morality.

This “great cultural revolution” is bound to be chaotic and to give rise to all kinds of grotesque developments. The disintegration of the old moralistic institutions and customs expresses itself at first as a rebellion that takes pathological forms, but it is not difficult to see that healthy forces are trying to break through in these pathological manifestations. At one time Reich envisaged a “powerful international organization” that would create an atmosphere of sex-affirmation and thus help to “guide the rebellion into rational channels” (*The Mass Psychology of Fascism*, p. 121). However, regardless of whether any organizations are brought into being which could accelerate the process and make it less painful, there is no reason to “fear for the final outcome.” As yet, human beings, “moved by obscure, ‘oceanic’ feelings, dream instead of mastering their existence; and they perish from these dreams” (*Character Analysis*, p. 324). But when once they master their existence, when they become capable of giving and receiving love and when work will be a source of pleasure and not a burden, this will mean “the death-knell of all transcendental mysticism, of the ‘absolute objective spirit,’” and of all the metaphysical and irrationalist philosophies that are “subsumed under mysticism in the ... wider sense.” An individual “who is sexually happy does not need an inhibiting ‘morality’ or a supernatural ‘religious experience.’ Basically, life is as simple as that. It becomes complicated only by the human structure which is characterized by the fear of life” (*The Sexual Revolution*, p. 269).

## REICH’S LAST YEARS

It is not surprising that the ideas sketched in the preceding sections of this entry should have appealed to many who were dissatisfied with the conservative developments of psychoanalysis as well as to those who, disillusioned with the results of communism in Russia, nevertheless strongly believed in social progress. During his early years in the United States, Reich did in fact count among his followers or sympathizers a number of remarkably talented men, from the most varied walks of life, who saw the dawn of a new enlightenment in his psychiatry and in the implications of his theories for education and for the proper direction of social reform. It would be difficult to convey to anybody who was not actually living in New York at that time the enthusiasm that was felt for Reich personally and for what were regarded as his liberating insights. As was to be expected, communists and psychiatrists of other schools were violently hostile, but this only served to heighten people’s admiration for Reich’s independence and for his uncompromising integrity.

It was mentioned previously that Reich himself became less and less interested in psychiatry. He also gradually lost most of his concern to guide into rational channels the “great cultural revolution” that he had diagnosed in his writings. The publications of his last years do indeed contain numerous discussions of social topics, but, at least in the opinion of the present writer, most of what Reich now had to say was flat and trivial. He became increasingly obsessed with the evil conspiracies of “red fascism” (some of Reich’s remarks during this period could be quoted with approval by members of the John Birch Society) and with the menace of the “emotional plague.” This term was originally introduced to refer to the harmful activities of individuals who take out their sexual sickness and frustrations on the rest of humankind, usually under the pretense of promoting some worthy cause.

Reich’s earlier description of emotional plague reactions and motives had been extremely perceptive, but now anybody who was in any way opposed to any of his ideas became automatically classified as an agent of the emotional plague. The writings of the last years are also filled, in a manner reminiscent of Friedrich Nietzsche’s *Ecce Homo*, with hymns of self-praise (sometimes in the third person), and there is much evidence of extreme bitterness toward a world that did not accept or even pay attention to his theories. From the available accounts it appears that Reich had always been impatient and somewhat autocratic, but he had also been singularly compassionate and generous. Dr. Nic Waal, in her sketch of

Reich, describes him as “enormously stimulating and lovable” but adds that in his last years he “became less and less patient, less loving ... and finally pathologically suspicious” (*Wilhelm Reich—A Memorial Volume*, p. 37).

If Reich became increasingly bitter, this was not without a good deal of justification. Right from the beginning, even while he was a psychoanalyst “in good standing,” Reich was the victim of an extraordinary amount of spite and slander. Any study of the records will make it clear that he was treated outrageously by the officials of the Psychoanalytic Association both before and at the Lucerne Conference. We have already mentioned Reich’s troubles in Scandinavia. In New York, he was arrested by the Federal Bureau of Investigation in December 1941 and held at Ellis Island for three weeks. The reasons for the arrest were never divulged. In 1947 an exceptionally vicious campaign was initiated in the *New Republic*, by the journalist Mildred E. Brady. There was not a paragraph in her article that did not contain a major distortion, but it was nevertheless quoted and reprinted all over America.

In an article ten years earlier, the German poet Stephan Lackner had expressed his indignation at the treatment that Reich had received and continued to receive from leading figures among the psychoanalysts and the left-wing parties. “It was not enough,” wrote Lackner, “to expel Reich from their organizations”; in the struggle against this man and his disturbing ideas, “every kind of slander and distortion is a permissible weapon” (*Das neue Tagebuch*, February 1937, p. 140). This last remark applies, word for word, to the campaign instigated by Brady and her associates. In March 1954, the U.S. Food and Drug Administration obtained an injunction against Reich and his foundation, ordering the destruction of all orgone accumulators, all of Reich’s journals, and some of his books; the books that were not destroyed were to be impounded. Among the works proscribed on the ground that they constituted “labeling” of the orgone accumulator were such books as *The Function of the Orgasm* and *Character Analysis*, in which the accumulator is not so much as mentioned.

Nobody except fanatical partisans of Reich can dispute the right of the Food and Drug Administration to intervene. When on the defensive, Reich denied that he had ever claimed any curative powers for the orgone accumulator, but the truth is that the literature is full of such claims. However, granting that the Food and Drug Administration had evidence to show the accumulator medically worthless (no such evidence has ever been published), the injunction is nevertheless a startling docu-

ment constituting a blanket attack on Reich’s character and his entire work.

Reich had two weeks in which to appeal, but to everybody’s consternation he refused to appear in court. Instead, he wrote a letter to the judge in the case, declaring that a court of law was not the appropriate place for adjudicating scientific questions. For some months Reich obeyed the injunction, but in October 1954 he notified the authorities that he was about to resume all the activities of his institute, including the sale of books and periodicals. This led to a trial in 1956, at which Reich was given the maximum sentence of two years in a federal penitentiary. Reich died of a heart attack eight months after he had started serving his sentence. All journals published by Reich’s institute that were seized by government agents were burned in two separate actions in 1956 and 1960, and his books were impounded until they began to be republished by a commercial house in 1960.

There is no doubt in the mind of the present writer that during his last years Reich was mentally ill. Some of those who were close to him deny this, and the prison psychiatrist who examined Reich certified him as sane. Nevertheless, if one reads the records of the trial or the brief that Reich filed in his appeal, one can hardly resist drawing the conclusion that a great man had broken down. Reich finally “went to pieces,” observed Dr. Waal, “partly on his own—but mostly due to other people,” adding that “a human being cannot bear cruelty and loneliness in the long run” (*Wilhelm Reich—A Memorial Volume*, pp. 38–39). It is worth recalling the words of Josef Popper-Lynkeus, whose ideas bear little resemblance to Reich’s but who was also described as “mad” during the better part of his life. “I assure you,” he told his biographer, “[that] of all experiences none is more painful than that of finding oneself described as mad as a consequence of having discovered something that is good and true: of all martyrdoms, this is perhaps the most terrible” (A. Gelber, *Josef Popper-Lynkeus*, p. 101).

**See also** Functionalism in Sociology.

### **Bibliography**

The following abbreviations are used throughout; ZPS for *Zeitschrift für politische Psychologie und Sexualökonomie* and IJSO for *International Journal for Sex-Economy and Orgone Research*.

Several biographies of Reich have been announced, but none had been published by the time this entry went to press. The only published sketches of Reich are A. S. Neill’s “The Man Reich” and Nic Waal’s “On Wilhelm Reich,” both in *Wilhelm Reich—A Memorial Volume*, edited by Paul Ritter (London, 1958). There is a good deal of autobiographical material,



especially on his relations with Freud, in Reich's *The Function of the Orgasm*, translated from the German manuscript by T. P. Wolfe (New York: Orgone Institute Press, 1942; paperback reprint, 1961). This book is a good introduction to all of Reich's theories discussed in the present entry. The reader should be warned, however, that in the 1961 reprint the very valuable introduction by Dr. Wolfe has been deleted. *People in Trouble* (Rangeley, ME: Orgone Institute Press, 1953) contains an account of Reich's work at his sex hygiene clinics and of his difficulties with communist functionaries in Germany and Denmark. Reich's attempt to organize an international movement in support of a sex-affirmative culture is described by him in two articles: "Zur Geschichte der Sexpol Bewegung," in *ZPS* 1 (1934): 259–269, and "Geschichte der deutschen Sexpol-Bewegung," in *ZPS* 2 (1935): 64–70. The only published account of Reich's troubles in Norway is Gunnar Leistikow, "The Fascist Newspaper Campaign in Norway," in *IJSO* 1 (1942): 266–273. This article also discusses Reich's troubles in Denmark. Its title is misleading in that many of Reich's opponents were not fascists. Reich's *Listen Little Man!*, translated from the German manuscript by T. P. Wolfe, with illustrations by William Steig (New York: Orgone Institute Press, 1948), is a moving outburst against the various people who harassed and defamed him.

The fullest published account of Reich's technique of vegetotherapy is *Orgasmusreflex, Muskelhaltung und Körperausdruck* (Copenhagen: Sexpol, 1937), parts of which are translated in Chapter 8 of *The Function of the Orgasm* and Chapter 15 of the third edition of *Character Analysis* (New York: Orgone Institute Press, 1949). Various aspects of Reich's new technique are also discussed in the following articles: Odd Havrevold, "Vegetotherapy," in *IJSO* 1 (1942): 65–87, written under the pseudonym Walter Frank; Ola Raknes, "The Treatment of a Depression," in *IJSO* 1 (1942): 163–170, and "Sex-Economy," in *IJSO* 3 (1944): 17–37, written under the pseudonym Carl Arnold. (These pseudonyms were necessary during the Nazi occupation of Norway.)

Child therapy is discussed in Felicia Saxe, "A Case History," in *IJSO* 4 (1945): 59–71, and "Armored Human Beings versus the Healthy Child," in *Annals of the Orgone Institute* 1 (1947): 35–72; and in Nic Waal, "A Case of Anxiety Neurosis in a Small Child," in *Bulletin of the Menninger Clinic* 12 (1948), and "A Special Technique of Psychotherapy with an Autistic Child," in *Emotional Problems of Early Childhood*, edited by G. Caplan (New York: Basic, 1955).

The fullest statement of Reich's views on religion is found in Chapters 6 and 7 of *Massenpsychologie des Faschismus* (Copenhagen: Verlag für Sexualpolitik, 1933), third edition translated by T. P. Wolfe as *The Mass Psychology of Fascism* (New York: Orgone Institute Press, 1946). Discussions of religion strongly influenced by Reich are J. H. Leunbach, "Religion und Sexualität," in *ZPS* 1 (1934): 70–72; Karl Teschitz, "Grundlagen der Religion," in *ZPS* 2 (1935): 103–129, and "Religiöse Ekstase als Ersatz der sexuellen Auslösung," in *ZPS* 4 (1937): 23–34; and Theodor Hartwig, "Religion und Sexualität," in *Der Freidenker* (Bern, April 1936), and "Der Sinn der religiös-sittlichen Erziehung," in *ZPS* 4 (1937): 203–205. Of philosophical interest are Reich's articles "Zur Anwendung der Psychoanalyse in der Geschichtsforschung," in *ZPS* 1 (1934): 4–16, and "Die Funktion der objektiven Wertwelt," in *ZPS* 2 (1935): 32–43.

For some years Reich considered himself a dialectical materialist. His attempt to give empirical meaning to the so-called dialectical laws can be found in *Dialektischer Materialismus und Psychoanalyse* (Berlin, 1929; 2nd ed., Copenhagen, 1934).

Reich's views concerning the relation between society and character structure are stated succinctly in "Charakter und Gesellschaft," in *ZPS* 3 (1936), translated by T. P. Wolfe as "Character and Society," in *IJSO* 1 (1942): 247–256, and much more fully in *The Mass Psychology of Fascism* and in *Die Sexualität im Kulturkampf* (Copenhagen, 1936), translated by T. P. Wolfe as *The Sexual Revolution* (New York: Orgone Institute Press, 1945). Reich's claims about the "function" of sexual suppression are partly based on his anthropological theories, which are an extension of the work of Malinowski. The fullest statement of these theories is found in *Der Einbruch der Sexualmoral* (Berlin: Verlag für Sexualpolitik, 1932). There is a critical discussion of Reich's anthropology in a review of this book by Erich Fromm, in *Zeitschrift für Sozialforschung* 2 (1933): 119–122.

Critical discussions of Reich's character-analytic technique are found in Carl M. Herold, "A Controversy about Technique," in *Psychoanalytic Quarterly* 8 (1939): 219–243; and in Richard Sterba, "Clinical and Therapeutic Aspects of Character Resistance," in *Psychoanalytic Quarterly* 22 (1953): 1–20. The dispute over the existence of "primary masochism" is surveyed in C. C. Flugel, *Man, Morals and Society* (London: Duckworth, 1945). Flugel, after some hesitation, sides with Theodor Reik and Franz Alexander against Reich. Even sympathetic commentators have frequently expressed doubts about what they take to be the excessively simple and "Rousseauist" view concerning the "natural" man that is implicit in many of Reich's discussions. Reich's view on this subject is condemned as "a stale left-over of the eighteenth-century imagination" in Philip Rieff, "The World of Wilhelm Reich," in *Commentary* 38 (September 1964): 50–58. There are replies to Rieff in *Commentary* 39 (February 1965): 19–22. Perhaps the best-known attack on Reich is found in Chapter 21 of Martin Gardner, *Fads and Fallacies in the Name of Science* (New York, 1952; 2nd ed., New York: Dover, 1957). In the opinion of the present writer, Gardner gives an extremely distorted picture of Reich's significance, concentrating on the wild claims of his last years and doing scant justice to the ideas discussed in the present entry.

Various of Reich's theories are sympathetically discussed in Max Hodann, *A History of Modern Morals* (London: Heinemann, 1937); Stephan Lackner, "Ein moderner Ketzer," in *Das neue Tagebuch* (Paris) 5 (1937): 140–141; Harald K. Schjelderup, *Nervose Og Opdragelse* (Oslo, 1937); Neil McInnes, "An Examination of the Work of Wilhelm Reich," in *Hermes* 48 (1946): 26–29; Paul Goodman, "The Political Meaning of Some Recent Revisions of Freud," in *Politics* 2 (1945): 197–203, and "Dr. Reich's Banned Books," in *Kulchur* (1960), reprinted in Goodman's *Utopian Essays and Practical Proposals* (New York: Vintage, 1964), C. Berg, *Psychotherapy* (New York, 1948); Rudolf Brun, *General Theory of Neuroses*, translated by B. Miall (New York, 1951); and R. A. Wilson, "Wilhelm Reich and the Book Burners," in *Minority of One* 3 (February 1961): 6–7. There are constant references to Reich's therapy and to his social theories in the books written by A. S. Neill from 1939 on, such as *The*

*Problem Teacher* (London: Jenkins, 1939), *Hearts Not Heads in the School* (London: Jenkins, 1945), and *The Problem Family* (New York: Hermitage Press, 1949).

The *Orgone Energy Bulletin* 5 (1953): 1–137, contains a very extensive bibliography of writings by and about Reich up to 1952. Unfortunately all issues of this periodical, as well as all issues of *IJSO*, are among the publications that were burned by the U.S. government.

#### OTHER RECOMMENDED WORKS

- Cohen, Ira H. *Ideology and Unconsciousness: Reich, Freud, and Marx*. New York: New York University Press, 1982.
- Grunberger, Béla, and Janine Chasseguet-Smirgel. *Freud or Reich?: Psychoanalysis and Illusion*. New Haven, CT: Yale University Press, 1986.
- Mann, W. E. *Orgone, Reich, and Eros: Wilhelm Reich's Theory of Life Energy*. New York: Simon and Schuster, 1973.
- Mann, W. E., and Edward Hoffman. *The Man Who Dreamed of Tomorrow: A Conceptual Biography of Wilhelm Reich*. Los Angeles: J.P. Tarcher; Boston: Houghton Mifflin, 1980.
- Reich, Ilse Ollendorff. *Wilhelm Reich: A Personal Biography*. New York: St. Martin's Press, 1969.
- Reich, Wilhelm. *American Odyssey: Letters and Journals, 1940–1947: Beyond Psychology*. New York: Farrar, Straus and Giroux, 1999.
- Reich, Wilhelm. *Beyond Psychology: Letters and Journals, 1934–1939*, edited by Mary Higgins. New York: Farrar, Straus, and Giroux, 1994.
- Reich, Wilhelm. *Character Analysis*. 3rd ed. New York: Farrar, Straus and Giroux, 1972.
- Reich, Wilhelm. *Early Writings*. New York: Farrar, Straus and Giroux, 1975.
- Reich, Wilhelm. *Passion of Youth: An Autobiography, 1897–1922*, edited by Mary Higgins and Chester M. Raphael. New York: Farrar, Straus, Giroux, 1988.
- Reich, Wilhelm. *Reich Speaks of Freud: Wilhelm Reich Discusses His Work and His Relationship with Sigmund Freud*. New York: Farrar, Straus and Giroux, 1967.
- Reich, Wilhelm. *Selected Writings: An Introduction to Orgonomy*. New York: Farrar, Straus and Giroux, 1973.
- Robinson, Paul A. *The Freudian Left: Wilhelm Reich, Geza Roheim, Herbert Marcuse*. New York: Harper & Row, 1969.
- Sharaf, Myron R. *Fury on Earth: A Biography of Wilhelm Reich*. New York: St Martin's Press/Marek, 1983.

**Paul Edwards (1967)**

*Bibliography updated by Michael J. Farmer (2005)*

## REICHENBACH, HANS

(1891–1953)

Hans Reichenbach was a leading philosopher of science and a proponent of logical positivism. He made important contributions to the theory of probability and to the philosophical interpretation of the theory of relativity, quantum mechanics, and thermodynamics.

## LIFE

Reichenbach studied civil engineering, physics, mathematics, and philosophy at Berlin, Göttingen, and Munich in the 1910s. Among his teachers were neo-Kantian philosopher Ernst Cassirer, mathematician David Hilbert, and physicists Max Planck, Max Born, and Arnold Sommerfeld. Reichenbach received his degree in philosophy from the Friedrich-Alexander University of Erlangen-Nürnberg in 1915 with a dissertation on the theory of probability titled *Der Begriff der Wahrscheinlichkeit für die mathematische Darstellung der Wirklichkeit* (The Concept of Probability for the mathematical Representation of Reality), published in 1916. Between 1917 and 1920, while he was working as a physicist and engineer, Reichenbach attended Albert Einstein's lectures on the theory of relativity at Berlin. He was fascinated by the theory of relativity and in a few years published four books about this subject: *The Theory of Relativity and A Priori Knowledge* (1920), *Axiomatization of the Theory of Relativity* (1924), *From Copernicus to Einstein* (1927), and *The Philosophy of Space and Time* (1928). In 1920 he began teaching at the Technische Hochschule at Stuttgart as private docent.

With the help of Einstein, Planck, and Max von Laue, in 1926 Reichenbach became assistant professor in the physics department of Berlin University. In 1930 he undertook the editorship of the journal *Erkenntnis* (Knowledge) with Rudolf Carnap. In 1933, soon after Adolf Hitler became chancellor of Germany, Reichenbach was dismissed from Berlin University because his family had Jewish origin. He emigrated to Turkey, where he was appointed chief of the philosophy department of Istanbul University with a five-year contract. During his stay in Turkey he published *The Theory of Probability* (1935). In 1938 he moved to the United States, where he became professor at the University of California at Los Angeles. In the following years Reichenbach published *Experience and Prediction* (1938); *Philosophic Foundations of Quantum Mechanics* (1944); *Elements of Symbolic Logic* (1947); "The Philosophical Significance of the Theory of Relativity" in *Albert Einstein: Philosopher-Scientist* (1949), edited by Paul Arthur Schilpp; and *The Rise of Scientific Philosophy* (1951). Reichenbach died in 1953 while he was working on the nature of scientific laws and on the philosophy of time. The two books that came from this work, *Nomological Statements and Admissible Operations* (1954) and *The Direction of Time* (1956), were published posthumously.

## COORDINATIVE DEFINITIONS

An important tool introduced by Reichenbach for the philosophical analysis of scientific theories is that of coordinative definitions. According to Reichenbach, a mathematical theory differs from a physical theory because the latter uses a specific type of definition, named coordinative definition, which coordinates (that is associates) some concepts of the theory with physical objects or processes. An example of a coordinative definition is the definition of the standard unit of length in the metric system, which connects the meter with a rod housed in the International Bureau of Weights and Measures in Sèvres, or with a well-defined multiple of the wavelength of a determined chemical element. Another example is the definition of the straight line as the path of a ray of light in vacuum. A scientific theory acquires a physical interpretation only by means of coordinative definitions. Without such type of definitions a theory lacks of a physical interpretation and it is not verifiable, but it is an abstract formal system, whose only requirement is axioms' consistency.

Geometry well illustrates the role of coordinative definitions. In Reichenbach's opinion, there are two different kinds of theories concerning geometry, namely mathematical geometry and physical geometry. Mathematical geometry is a formal system that does not deal with the truth of axioms, but with the proof of theorems—that is, it only searches for the consequences of axioms. Physical geometry is concerned with the real geometry in the physical world; it searches for the truth or falsity of axioms using the methods of the empirical science. The physical geometry derives from the mathematical geometry when appropriate coordinative definitions are added. For example, if the concept of a straight line is coordinated with the path of a ray of light in a vacuum, the theory of relativity shows that the real geometry is a non-Euclidean geometry. Without coordinative definitions, Euclidean and non-Euclidean geometry are nothing but formal systems; with coordinative definitions, they are empirically testable. Coordinative definitions are conventions, because it is admissible to choose a different definition for a concept of a theory. In the case of geometry, with a different definition for the straight line, Euclidean geometry is true. In a sense, choosing between Euclidean and non-Euclidean geometry is not a matter of facts, but a matter of convention.

## RELATIVITY OF GEOMETRY

Reichenbach insists on the importance of the coordinative definitions in his philosophical analysis of the theory

of relativity, especially in connection with the problem of determining the geometry of this world. In principle, scientists can discriminate between different geometry by means of measurements. For example, on the surface of a sphere, the ratio of the circumference of a circle to its diameter is less than  $\pi$ , whereas on the surface of a plane this ratio is equal to  $\pi$ . With a simple measurement of a circumference and of its diameter, we can discover we live on a sphere (the surface of Earth) and not on a plane. In the same way, using more subtle measurements, scientists can discover we live in a non-Euclidean space. However, there is a fundamental question: is measuring a matter of facts or does it depend on definitions? Reichenbach proposes the following problem, discussed in *The Philosophy of Space and Time*: is the length of a rod altered when the rod is moved from one point of space, say A, to another point, say B? We know many circumstances in which the length is altered. For example, the temperature in A can differ from the temperature in B. However, the temperature acts in a different way on different substances. If the temperature is different in A and in B, then two rods of different material, such as wood and steel, which have the same length in A, will have a different length in B. So we can recognize a difference in temperature and use suitable procedures to eliminate variations in measurement due to variations in temperature. In general, this is also possible for every differential force—that is, for every force that acts in a different way on different substances. But there is also another type of forces, called universal forces, which produce the same effect on all types of matter.

The best-known universal force is gravity, whose effect is the same on all bodies. What happens if a universal force alters the length of all rods, in the same way, when they are moved from A to B? By the very definition of universal forces, there are no observable effects. If we do not exclude universal forces, we cannot know whether the length of two measuring rods, which are equal when they are in the same point of space, is the same when the two rods are in two different points of space. Excluding universal forces is nothing but a coordinative definition. We can also adopt a different definition, in which the length of a rod depends on the point of space in which the rod stays. So the result of a measurement depends on the coordinative definition we choose. As a consequence, the geometrical form of a body, which depends on the result of measurements, is a matter of definition. The most important philosophical consequence of this analysis concerns the relativity of geometry. If a set of measurements supports a geometry G, we can arbitrarily choose a different geometry G' and adopt a different set of coordinative definitions so that the same set of meas-

urements supports  $G'$ , too. This is the principle of relativity of geometry, which states that all geometrical systems are equivalent. According to Reichenbach, it falsifies the alleged a priori character of Euclidean geometry and thus falsifies the Kantian philosophy of space.

### CAUSAL ANOMALIES

The principle of relativity of geometry is true for metric relationships—that is, for geometric properties of bodies depending on the measurement of distances, angles, and areas. The situation seems different when we are concerned about topology, which deals with the order of space—that is, the way in which the points of space are placed in relation to one another. A typical topological relationship is “point A is between points B and C.” The surface of a sphere and the surface of a plane are equivalent with respect to metrics, provided an appropriate choice of the coordinative definitions, but they differ from a topological point of view.

Consider the following example presented by Reichenbach in *The Philosophy of Space and Time*. Intelligent beings living on the surface of a sphere can adopt coordinative definitions that, from a metric point of view, transform the surface of the sphere into the surface of a plane. However, there is an additional difficulty: Because the surface of a sphere is finite, it is possible to do a round-the-world tour, walking along a straight line from a point A and eventually returning to the point A itself. Of course this is impossible on a plane, and thus it would seem that these intelligent beings have to abandon their original idea that they are living on a plane and instead must recognize they are on a sphere. But this is not true, because another explanation is possible: They can assert that they had walked in a straight line to point B, which is different from point A but, in all other respects, is identical to A. They can also fabricate a fictitious theory of pre-established harmony—according to which everything that occurs in A immediately occurs in B—in order to explain the similarity between A and B. This last possibility entails an anomaly in the law of causality. We can reject causal anomalies, but only by means of an arbitrary definition. Thus topology depends on coordinative definitions, and the principle of relativity of geometry also holds for topology. According to Reichenbach, this example is another falsification of Kantian theory of synthetic a priori. In Kantian philosophy, the Euclidean geometry and the law of causality are both a priori, but if Euclidean geometry is an a priori truth, normal causality can be false; if normal causality is an a priori truth, Euclidean geometry can be false. We arbitrarily can choose the

geometry, or we arbitrarily can choose the causality, but we cannot choose both.

### QUANTUM MECHANICS

Quantum mechanics differs from the other scientific theories because in this theory there is no possibility to introduce normal causality. No set of coordinative definitions can give an exhaustive interpretation of quantum mechanics free from causal anomalies.

It is important to explain some concepts used by Reichenbach in *Philosophical Foundations of Quantum Mechanics*, his main work about quantum mechanics. Using a wider sense of the word “observable,” some events occurring in quantum mechanics are observable; they are events consisting in coincidences between particles or between particles and macroscopic material, like the collision of an electron on a screen, signaled by a flash of light. Events between such types of coincidences are unobservable; an example is the path of an electron between the source and the screen on which it collides.

Quantum observable events are called, by Reichenbach, phenomena, whereas unobservable ones are called interphenomena. Reichenbach explains that there are three main interpretations concerning interphenomena: wave interpretation, according to which matter consists of waves; corpuscular interpretation, according to which matter consists of particles; and Bohr-Heisenberg interpretation, according to which statements about interphenomena are meaningless. The first two interpretations are called exhaustive interpretations, because they include a complete description of interphenomena. The last is a restricted interpretation, because it prohibits assertions about interphenomena. A normal system is an interpretation in which the laws of nature are the same for phenomena and interphenomena. This definition of a normal system is modeled on a basic property of classical physics: the laws of nature are the same whether or not the object is observed.

With these definitions, it is possible to formulate Reichenbach’s principle of anomaly in quantum mechanics: there is no normal system. Thus causal anomalies cannot be removed from quantum mechanics. However, there is another peculiarity in quantum mechanics: for every experiment there exists an exhaustive interpretation—which is a wave or a corpuscular interpretation—that provides a normal system, although limited to this experiment. In other words, there does not exist an interpretation free from all causal anomalies, but for every causal anomaly there does exist an interpretation that ruled out this anomaly. For example, if we adopt the cor-

puscular interpretation, we have to face causal anomalies raising from some experiments, such as the two-slits experiment. In this experiment a beam of electrons is directed toward a diaphragm with two open slits and an interference pattern is produced on a screen behind the diaphragm; the probability that an electron, passing through an open slit, will reach the screen at a given point is depending on whether the other slit is open or closed—with the electron behaving as if it is informed about the state of the other slit.

This causal anomaly is eliminated if we adopt the wave interpretation, according to which the interference patterns are produced by waves in conformity with Huygens's principle. The wave interpretation is in turn affected by other anomalies raising from the so-called reduction of the wave packet: The wave originating from an open slit occupies a hemisphere centered on the slit, but when the wave hits the screen, a flash is produced in a point only and the wave disappears in all other points. Apparently all physical properties transported by the wave, such as momentum and energy, suddenly materialized in a single point, even if they were distant from this point. This situation is explained without anomalies by the corpuscular interpretation. According to Reichenbach, in every experiment about quantum mechanics we can adopt an interpretation free from causal anomalies, but we have to use a different interpretation in a different experiment. Only two interpretations are required: the wave and the corpuscular interpretation. This is the real meaning of the duality of wave and corpuscle in quantum physics. The possibility of eliminating causal anomalies from every quantum experiment is called, by Reichenbach, the principle of eliminability of causal anomalies.

The Bohr-Heisenberg restricted interpretation of interphenomena named after Danish physicist Niels Bohr and German physicist Werner Karl Heisenberg, states that speaking about values of unmeasured physical quantities is meaningless. Reichenbach criticizes the Bohr-Heisenberg interpretation on two points. First, Heisenberg's indeterminacy principle becomes a meta-statement about the semantics of the language of physics; second, this interpretation implies the presence of meaningless statements in the language of physics.

Using a three-valued logic, in which admissible truth values are *truth*, *falsehood*, and *indeterminacy*, Reichenbach constructs another restrictive interpretation in which a statement about an unmeasured physical quantity can be neither true nor false, but indeterminate.

## INTERPRETATIONS OF REICHENBACH'S PHILOSOPHY

An open question regards the relation between Reichenbach and conventionalism. His insistence on the major role played by the coordinative definitions, the relativity of geometry, the equivalence between wave and corpuscular interpretation of quantum mechanics has suggested that his philosophy can be ascribed to conventionalism. In Reichenbach's works there are some points corroborating this view. For example, he asserts that the philosophical meaning of the theory of relativity is that this theory proves the necessity of coordinative definitions, which are arbitrary, in situations in which empirical relations had been previously assumed. But there are also some elements against the conventionalist reading of Reichenbach's philosophy, as seen in the last paragraph of *The Philosophy of Space and Time*, in which Reichenbach affirms that the reality of space and time is an irrefutable consequence of his epistemological analysis; it is an assertion apparently incompatible with conventionalism. As an example of the debate about Reichenbach's attitude toward conventionalism, it is possible to mention the conventionalist interpretation of Reichenbach's philosophy developed by Adolf Grünbaum in *Philosophical Problems of Space and Time* (1973) and Hilary Putnam's counterarguments offered in "The Refutation of Conventionalism" (1975).

A different explication of Reichenbach's philosophy, based on an analysis of the role of the coordinative definitions in the light of Kantian philosophy, is advanced by Michael Friedman and exposed in *Reconsidering Logical Positivism* (1999). According to Friedman's interpretation, Reichenbach, in his first published work on the theory of relativity (*Theory of Relativity and A Priori Knowledge*), distinguishes two different meanings of synthetic a priori, which are united in Kantian philosophy. In the first meaning, a synthetic a priori judgment is necessary and thus not modifiable; in the second meaning, a synthetic a priori statement is constitutive of the object. The coordinative definitions are not necessary judgments, because we can make use of a different definition. Moreover, all coordinative definitions are subjected to changes with the evolution of knowledge, so they are modifiable. Thus they are not a priori in the first meaning present in Kantian philosophy. But the coordinative definitions are required to give an empirical interpretation to a theory and so they are constitutive of the object of knowledge. Thus they are synthetic a priori in the second meaning present in Kantian philosophy. Friedman calls this type of a priori judgment "constitutive, rela-

tivized a priori” (1999, p. 62), because they are a priori in the constitutive sense, relative to a given theory.

Surely Kantian philosophy exerts a great influence on Reichenbach. He professes admiration for Kant in his first works. In the article “Kant und die Naturwissenschaft” (1933, p. 626) he says, “There is no doubt that he [Kant] was one of the few thinkers whose work showed the way on which the contemporary philosophy of natural science continues to proceed.” According to Reichenbach, Kantian philosophy of nature is a meaningful theory, although it is superseded by the outcomes of contemporary physics. Later, Reichenbach accentuates his departure from Kant, stressing his criticism of synthetic a priori and developing many arguments against Kantian philosophy.

**See also** Causation: Philosophy of Science; Philosophy of Statistical Mechanics; Time.

### Bibliography

- Barone, Francesco. *Il neopositivismo logico*. Bari, Italy: Laterza, 1986.
- Friedman, Michael. *Dynamics of Reason*. Stanford, CA: Center for the Study of Language and Information, 2001.
- Friedman, Michael. *Reconsidering Logical Positivism*. New York: Cambridge University Press, 1999.
- Grünbaum, Adolf. *Philosophical Problems of Space and Time*. Dordrecht, Netherlands: Reidel, 1973.
- Putnam, Hilary. “The Logic of Quantum Mechanics.” In *Mathematics, Matter and Method. Philosophical Papers*. Vol. 1. Cambridge, U.K.: Cambridge University Press, 1975.
- Putnam, Hilary. “The Refutation of Conventionalism.” In *Mind, Language and Reality. Philosophical Papers*. Vol. 2. Cambridge, U.K.: Cambridge University Press, 1975.
- Salmon, C. Wesley, and Gereon Wolters, eds. *Logic, Language, and the Structure of Scientific Theories: Proceedings of the Carnap-Reichenbach Centennial, University of Konstanz, 21–24 May 1991*. Pittsburgh: University of Pittsburgh Press, 1994.
- Schilpp, Paul Arthur, ed. *Albert Einstein: Philosopher-Scientist*. Evanston, IL: Library of Living Philosophers, 1949.
- Spohn, Wolfgang, ed. *Erkenntnis Orientated: A Centennial Volume for Rudolf Carnap and Hans Reichenbach*. Boston: Kluwer Academic Publishers, 1991.
- REICHENBACH’S WORKS
- “Kant und die Naturwissenschaft.” *Die Naturwissenschaften* 21 (1933): 601–606, 624–626.
- From Copernicus to Einstein*. New York: Alliance Book Corp., 1942.
- Philosophic Foundations of Quantum Mechanics*. Berkeley: University of California Press, 1944.
- Elements of Symbolic Logic*. New York: Macmillan, 1947.

*The Theory of Probability: An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability*. Translated by Ernest H. Hutten and Maria Reichenbach. Berkeley: University of California Press, 1949.

*The Rise of Scientific Philosophy*. Berkeley: University of California Press, 1951.

*The Direction of Time*. Edited by Maria Reichenbach. Berkeley: University of California Press, 1956.

*The Philosophy of Space and Time*. Translated by Maria Reichenbach and John Freund. New York: Dover Publications, 1957.

*Axiomatization of the Theory of Relativity*. Translated and edited by Maria Reichenbach. Berkeley: University of California Press, 1969.

*Laws, Modalities, and Counterfactuals*. Berkeley: University of California Press, 1976.

**Mauro Murzi (2005)**

## REID, THOMAS

(1710–1796)

Thomas Reid was the founder of the Scottish “Common Sense” school of philosophy. A contemporary and critic of David Hume, he is best known for his staunch defense of common sense and trenchant opposition to the “way of ideas,” the theory that the immediate objects of perception and other cognitive acts are always internal images or ideas, not external physical objects. His views exerted a good deal of influence until the mid-nineteenth century or so, when they began to be eclipsed by absolute idealism, pragmatism, and other philosophical movements, but they have been the subject of renewed interest from the 1970s on.

After being educated at Marischal College in Aberdeen, Scotland, Reid served for fifteen years as a parish minister in nearby New Machar. In 1752 he was appointed professor at King’s College in Aberdeen, where he taught mathematics, physics, and philosophy, among other subjects. He tells us that in his youth he believed nearly the entire philosophy of George Berkeley but that a reading of Hume’s *Treatise of Human Nature* (1739–1740) convinced him (by carrying Berkeley’s philosophy to its logical conclusions) that there must be some original defect in it. Reid identified this defect as the theory of ideas, which he went on to challenge in college lectures, meetings of the Aberdeen Philosophical Society, and two books. In 1764 he published his first major work, *An Inquiry into the Human Mind on the Principles of Common Sense*, in which he set forth his reasons for opposing the theory of ideas and offered an alternative theory of how we gain knowledge by means of the various senses.

In the same year he accepted the chair in moral philosophy at Glasgow, succeeding Adam Smith. He lectured there until 1780, when he resigned to prepare his last two major works: *Essays on the Intellectual Powers of Man* (1785), devoted to the contributions of perception, memory, reason, and other cognitive powers to human knowledge, and *Essays on the Active Powers of the Human Mind* (1788), devoted to the nature of action, will, freedom, and morality.

This article provides a synopsis of Reid's main views more or less in the order in which he presented them in his three published books: the *Inquiry* (abbreviated as Inq.), the *Intellectual Powers* (abbreviated as IP), and the *Active Powers* (abbreviated as AP). Numbers separated by a period refer to chapter and section numbers in the *Inquiry* and to essay and chapter numbers in the two volumes of *Essays*.

### CRITIQUE OF THE THEORY OF IDEAS

Almost alone among the great modern philosophers, Reid espoused a direct realist theory of perception. He repudiated the assumption that what is immediately present to the mind is never an external thing, but only an internal image, impression, representation, or (to use the most common eighteenth-century term) *idea*. Ideas were usually conceived of as mental entities that existed only as long as the mind was aware of them. Reid found the theory of ideas to be taken for granted in the work of most of his philosophical predecessors, including René Descartes, Nicolas Malebranche, Antoine Arnauld, John Locke, Berkeley, and Hume. Some of these philosophers (for example, Descartes and Locke) were realists, believing that ideas are caused in us by physical objects existing outside the mind. Others (notably, Berkeley) were idealists, repudiating the existence of a world outside the mind and believing that the things we call physical objects are simply bundles of ideas. In either case, we are cut off from direct perception of the physical world, either because there is no physical world to be perceived, or because our perception of it is indirect—not strictly perception at all, but inference based on what we do perceive, namely, ideas.

Reid makes at least three important points against the theory of ideas. First, the arguments in favor of the theory are weak and without cogency; second, the theory does nothing to explain how perception is possible; third, the theory stands in the way of our knowing or even being able to conceive of the physical world.

One of the arguments for the theory of ideas that Reid singles out for criticism is a version of the argument

from perceptual relativity. Hume had claimed that the “universal and primary opinion of all men” that they perceive external objects directly is “destroyed by the slightest philosophy,” offering the following argument in section 12 of the *Enquiry concerning Human Understanding*: “The table, which we see, seems to diminish as we remove further from it; but the real table, which exists independent of us, suffers no alteration. It was therefore nothing but its image which was present to the mind.” Hume's argument may be cast into the following syllogism: (1) What I see diminishes in magnitude as I retreat from it; (2) the table itself does not diminish in magnitude as I retreat from it; (3) therefore, what I see is not the table itself (but only an image or idea).

Reid contends that Hume's premises are true only if we restate them as follows (IP 2.14, p. 182): (1) What I see diminishes in *apparent* magnitude as I retreat from it; (2) the table itself does not diminish in *real* magnitude as I retreat from it; (3) therefore, what I see is not the table (but only an image or idea).

The real magnitude of an object (for example, the edge of a table) is an intrinsic property of it, measured in feet or inches, whereas the apparent magnitude of an object is a relation between the object and a perceiver (or his vantage point), measured by the angle the object subtends at the eye. Reid takes the terminology of “real” versus “apparent” from the astronomy of his day; it is not necessarily implied that there is anything illusory about apparent magnitude. It is easy to see that apparent magnitude varies with the distance between object and perceiver (objects subtending smaller angles when further away) whereas real magnitude does not. Once we record these facts correctly, as in Reid's version of the syllogism, we see that the conclusion of the argument does not follow from the premises. Moreover, Reid would resist the thought that if O has greater apparent magnitude when seen from p than when seen from p', that is because it presents a larger image to the observer at p than to the observer at p'. Apparent magnitude is a strictly dyadic relation, involving only the object and the perceiver (or his vantage point) and no third thing such as a mental image.

Reid's second point against the hypothesis of ideas is “that ideas do not make any of the operations of the mind to be better understood” (p. 184). Ideas had been thought necessary to explain how we perceive things that are distant, remember things that are past, or imagine things that do not exist at all, but Reid argues that all such explanations are worthless. They presuppose that ideas themselves can somehow be of the remote, the past, or the

nonexistent. But if ideas can do that, what prevents the simple idealess acts of perceiving, remembering, and imagining from doing it as well? Moreover, our ability to be aware of ideas themselves is no less mysterious than our ability to be aware of things that are not ideas.

It is as difficult to conceive how the mind perceives images in the brain as how it perceives things more distant. If any man will shew how the mind may perceive images in the brain, I will undertake to shew how it may perceive the most distant objects: for if we give eyes to the mind, to perceive what is transacted at home in its dark chamber, why may we not make these eyes a little longer-sighted? (Inq. 6.12, p. 121)

Reid's third point against the theory of ideas is "that the natural and necessary consequences of it furnish a just prejudice against it to every man who pays a due regard to the common sense of mankind" (p. 185). Chief among these consequences is that if we do not simply see or touch external objects, it becomes necessary to prove their existence by arguments. Descartes, Malebranche, and Locke all tried to muster such arguments, but none of the arguments is convincing. Reid thus thinks that skepticism about the material world is a built-in consequence of the theory of ideas. By contrast, if what we see and touch are not ideas but things in the external world, as in Reid's own view, this source of skepticism is eliminated.

## SENSATION AND PERCEPTION

A sensation is an event that occurs in a sentient subject when he or she smells a rose or tastes a fig. It lacks figure and extension and other qualities of bodies, being entirely mental. Reid calls sensations "principles of belief," by which he means that when we have a sensation and attend to it, we cannot help believing that it exists, that a subject of it exists (ourselves), and that some external object (for example, some quality in the rose) exists as its cause.

Reid is among the first to distinguish between sensation and perception. He explains this distinction as follows:

Thus, *I feel a pain*; *I see a tree*: the first denoteth a sensation, the last a perception. The grammatical analysis of both expressions is the same: for both consist of an active verb and an object. But, if we attend to the things signified by these expressions, we shall find, that in the first, the distinction between the act and the object is not real but grammatical; in the second, the distinction is not only grammatical but real. The form

of the expression, *I feel pain*, might seem to imply, that the feeling is something distinct from the pain felt; yet, in reality, there is no distinction. As *thinking a thought* is an expression which could signify no more than *thinking*, so *feeling a pain* signifies no more than *being pained*. What we have said of pain is applicable to every other mere sensation. (Inq. 6.20, pp. 167–68)

When I perceive a tree, there is an object (the tree) apart from my act of seeing, but when I have a sensation, there is no object apart from the act of sensing. As he defines sensation in the *Intellectual Powers*, it is an act of the mind "which may be distinguished from all others by this, that it hath no object distinct from itself" (IP 1.1, p. 36). That formulation is ambiguous: Does an act of sensing have itself for its object, or does it have no object at all? Although Reid's language often suggests the former option, his proposal that *being pained* is the model for all sensation suggests the latter option. If we take Reid in the latter way, he is a precursor of "adverbial" theories of sensation, such as were developed by C. J. Ducasse and Roderick Chisholm two centuries later: to have a sensation of red is not to sense something, but is simply to sense somehow—"redly," as the adverbial theory styles it.

Some critics of Reid have thought that his sensations are simply ideas under a new name, but there are important differences—especially if he holds an adverbial theory rather than a theory that divides sensation into act and object. If sensing required its own special objects, the argument from perceptual relativity against direct realism could be reinstated. The mountain that looks blue from a distance and green from close up would do so by generating first blue and then green sensory objects in my mind, and these variously colored objects would have to be distinct from the unchanging mountain. They would displace the mountain itself as my object of direct awareness. But Reidian sensations do not have objects to get in the way of direct perception of external things.

Although sensations do not *have* objects, they can *become* objects for us, in the sense that we can know through proper attention what sorts of sensations we are having. Indeed, Reid thinks that if we attend carefully to our sensations, we can know perfectly what they are like and can scarcely make any mistake about them. Yet typically we pay so little attention to them that we become almost oblivious to them; they serve as mere cues or signs from which our minds leap instantly to other things that they signify.



Our apprehension of that which sensations signify is *perception*. Reid's official characterization of perception involves three elements: conception, belief, and immediacy:

If, therefore, we attend to that act of our mind which we call the perception of an external object of sense, we shall find in it these three things:—*First*, Some conception or notion of the object perceived; *Secondly*, A strong and irresistible conviction of its present existence; and *Thirdly*, That this conviction and belief are immediate, and not the effect of reasoning. (IP 2.5, p. 96; cf. Inq. 6.20, p. 168)

Note that this definition makes no mention of sensation. Although Reid says that sensation generally serves as the trigger for the conception and belief involved in perception, perception proper is just conception plus immediate belief. Reid thinks it possible that perception should occur in the absence of sensation, and he holds that there is one variety of human perception that actually does occur without any characteristic sensation: namely, the perception of visible figure. Reid thus deemphasizes the role of sensation in perception in a way that some contemporary theorists (for example, James J. Gibson) would applaud. By the same token, however, his threefold definition may strike others as leaving out precisely that by which a genuine perception of a snake in the path ahead is distinguished from the conception and immediate belief in it one may form as the result of a friend's warning. Here Reid's views may gain in plausibility if we reckon his "conception" as something like what Bertrand Russell called knowledge by acquaintance. It is not necessarily the exercise of a concept in mere thought.

## REID'S NATIVISM

Reid thought that much of what he found alarming in Hume's philosophy stemmed from Hume's adherence to the empiricist maxim that we have no ideas or notions that are not derived from previous impressions or sensations. It is by this principle that Hume was led to conclude that we have no legitimate notions of objects existing unperceived, of causal connections amounting to more than constant conjunctions, and of a self that is the subject of various mental operations. Reid sought to overthrow Hume's philosophy by undermining its foundations, and for this purpose he tackled the empiricist principle head-on. He pointed to a notion that he thought Hume would surely concede that we possess—the notion of *extension*, or being spread out in space—and contended that this notion lacks a proper Humean

birthright in our sensations. If it were once acknowledged that not even so uncontroversial a notion as extension can be extracted from our sensations or impressions, Reid thought, the way would be open for recognizing the legitimacy of other notions with no sensory origin, such as the ideas of agency, self, and an external world.

To back up his contention that the notion of extension is not derived from sensation, Reid offers a thought experiment he calls his *experimentum crucis* (Inq. 5.6 and 5.7, pp. 65–72). He asks us to imagine a being furnished with a progressively richer array of sensations, beginning with those caused by the prick of a pin, advancing to more complex sensations such as those caused by the pressure of a blunt object against his or her body, and culminating with the sensations accompanying the motion of his or her limbs. He asks at each step in the series whether those sensory materials would suffice to give a being who reflected upon them a conception of extension, and his answer is no. Positively, Reid's doctrine is that the conception of extension is innate—not in the sense that we have it from birth, but in the sense that it is triggered in us by certain sensations from which it could never have been derived from any process of abstraction or ratiocination. We are enabled to form the conception of extended things only because we are innately programmed to do so.

For further light on the import of Reid's nativism, we may restate it in terms of the threefold classification of natural signs he offers in sections 4.2 and 5.3 of the *Inquiry*. Reid first divides signs into the artificial and the natural. In the former class, the connection between sign and thing signified is established by compact or convention, as with the words of human language. In the latter class, the connection between sign and thing signified is established by nature, as with smoke and fire and other cases of effect and cause. Reid then further divides natural signs into three classes. In the first class, the connection between sign and thing signified is "established by nature, but discovered only by experience" (Inq. 5.3, p. 59), as in the example of smoke and fire already given. In the second class, the connection is "not only established by nature, but discovered to us by a natural principle, without reasoning or experience" (Inq. 5.3, p. 60).

Reid thinks that certain features of the human countenance are signs in this sense of thoughts and other mental states. For example, an infant is innately disposed to read a smile on its mother's face as a sign of approval without having to learn this connection through experience. Unless there were a basic repertoire of natural signs of this second class, Reid believes, the signification of arti-

ficial signs could never be agreed upon or learned. Finally, in the third class are those signs “which, though we never before had any notion or conception of the thing signified, do suggest it, or conjure it up, as it were, by a natural kind of magic” (Inq. 5.3, p. 60). Not only is the *connection* between sign and thing signified innately programmed into our constitution (as with signs of the second class), but also the very *notion* of the thing signified is innate in the sense that it is in no way derivable by abstraction from any of our sensations. Reid believes that the tactile sensations to which we respond with conceptions of extended bodies are natural signs belonging to this third class.

Reid takes his nativism to afford an answer to an argument for skepticism he finds embodied in the combined philosophies of Berkeley and Hume. He formulates the argument as follows (Inq 5.8, p. 75): (1) We can have no conception of anything but what either resembles or is deducible from our sensations; (2) nothing resembles or is deducible from sensations but other sensations; (3) therefore, we can have no conception of anything but sensations.

If the argument were correct in both its premises, it would follow that we cannot even *conceive* of, let alone have knowledge of, a world lying beyond our sensations. Reid thinks the second premise is correct, and he credits Berkeley with having made it evident. But he thinks the first premise—which states in Reid’s language Hume’s principle that all our ideas are copied from precedent impressions—is false. “That we have clear and distinct conceptions of extension, figure, motion, and other attributes of body, which are neither sensations, nor like any sensation, is a fact of which we may be as certain, as that we have sensations” (Inq. 5.8, p. 76).

## THE MECHANICS AND GEOMETRY OF VISION

More than half of the *Inquiry* is devoted to vision, which Reid regards as the noblest of the senses. It informs us of the properties of objects far distant, such as the sun and the moon, and it can disclose in a glance the figure of a cathedral, whose delineation by touch would be the work of a lifetime.

Reid provided solutions to a number of puzzles about vision that lie today within the province of cognitive science rather than philosophy. For example, why do we see things upright despite having inverted retinal images of them? To explain this, Reid appeals to the law that an object will be seen in the direction of a straight line drawn from the point of retinal stimulation through

the center of the eye and into ambient space. Why do we normally see objects single despite having two retinal images of them, yet under certain circumstances see them double? Reid’s answer appeals to the law of corresponding retinal points: If rays from an object fall on points of the two retinas lying at equal distances and directions from their centers, the object will be seen as single, but otherwise as double.

One of Reid’s more remarkable findings is that the visible figures of objects are governed by a non-Euclidean geometry. Reid believed that sight by itself (before we have learned any correlations with touch) informs us only of the two-dimensional spatial features of objects. Although the objects we see are at a distance from us (*pace* Berkeley), the eye is incapable of making any discriminations of depth. To an eye placed at the center of a sphere and looking out, great circles on the surface of the sphere (whose outward curvature is invisible to the eye) must appear as straight lines, and every figure seen by the eye must have the same geometrical properties as some figure drawn upon the sphere. In consequence of this, Reid argued that the geometry of visibles is what we would nowadays classify as a Riemannian geometry. A visible triangle, unlike a triangle perceived by touch, always has an angle sum at least slightly greater than 180 degrees, and no two visibly straight lines are ever strictly parallel.

## MEMORY

Essay III of the *Intellectual Powers* is devoted to memory. Reid characterizes memory as “an immediate knowledge of things past” (IP 3.1, p. 253). There are two senses in which this is true. First, the object of memory is the very thing or event formerly perceived, not some present idea or simulacrum of it (Inq. 2.3, p. 28). Second, the knowledge one has by memory of this past object is noninferential; it does not rest on any reasoning from premises. Memory is thus like perception for Reid in involving both the conception of an object and an immediate belief in its existence; but it differs from perception because the object is, and is believed to be, past. Reid is severely critical of Hume’s attempt to distinguish imagining, remembering, and perceiving solely in terms of the force and vivacity of their objects.

Reid criticizes Locke’s view that memory is what constitutes personal identity—that person A is identical with person B who existed in the past if and only if A remembers what B did. He insists that memory is the *evidence* of personal identity, rather than that in which it consists (IP 3.4, p. 265). He also presents the famous

“brave officer” objection to Locke’s theory, courtesy of his friend George Campbell: Suppose a man who has become a general late in life remembers capturing the enemy flag as a young officer; that as an officer he remembered being flogged as a boy for robbing an orchard; and that as a general he no longer remembers being flogged as a boy. It follows from Locke’s theory that the general both is and is not the same person as the boy (IP 3.6, p. 276).

## CONCEPTION AND ABSTRACTION

Essays IV and V of the *Intellectual Powers* are devoted to conception and abstraction. Conception, the most basic operation of the mind, is “that operation of the understanding which the logicians call *simple apprehension*,” that is, the apprehension of a thing without any belief or judgment about it (IP 4.1, p. 295). “Judgment can be expressed by a proposition only, and a proposition is a complete sentence; but simple apprehension may be expressed by a word or words, which make no complete sentence” (IP 6.1, p. 408). The objects of conception expressed by words or subsentential phrases are either individuals or universals. What Reid calls conception should not be confused with conceptualization—that is, subsuming something under a concept—for the latter is judgment, and conception is more basic than judgment.

Reid holds that all the operations of our minds except sensation have objects distinct from themselves. “He that conceives, must conceive something” (IP 4.1, p. 311), and the same goes for perception and memory. It is a distinctive feature of conception, however, that its objects need not exist: “it is not employed solely about things which have existence” (IP 4.1, p. 310).

On this point, Reid is sometimes seen as a precursor of Alexius Meinong, who held that there can be cognitive relations to the utterly nonexistent and that a thing therefore need not exist in order to stand in relations. Meinong’s view strikes many as paradoxical. Yet Reid makes it look like one more piece of common sense or, at any rate, a consequence of two pieces of common sense (IP 4.1, p. 311): (1) I can conceive of a centaur; (2) no centaur exists; (3) therefore, I can conceive of what does not exist. In case anyone objects that the truth in premise 1 is simply that I can conceive of the *idea* of a centaur, which *does* exist, Reid is ready with a reply: I know the difference between conceiving of a centaur and conceiving of the idea of a centaur, and I can assure you that I am doing the former rather than the latter (IP 4.2, p. 321).

Reid’s view that the objects of conception may be nonexistent has an interesting application to the problem of abstract ideas, which pitted Locke against Berkeley and

Hume. On this topic, Reid writes, “Mr. Locke and his two antagonists have divided the truth between them” (IP 5.6, p. 394). Locke saw clearly “that the power of forming abstract and general conceptions is one of the most distinguishing powers of the human mind,” but he did not see “that this power is perfectly irreconcilable to his doctrine concerning ideas.” Berkeley and Hume “saw this inconsistency; but instead of rejecting the hypothesis of ideas, they explain away the power of abstraction.”

To see how Locke and his critics “divided the truth between them,” consider the following inconsistent triad of propositions:

- (1) We are sometimes aware in thought of the general and the abstract—in Reid’s terminology, we have the power of forming abstract and general conceptions.
- (2) We can only be aware of what exists: “in all of the operations of the understanding, there must be an object of thought, which really exists while we think of it” (IP 4.2, p. 312).
- (3) General entities have no existence: “every thing that really exists is an individual” (IP 5.6, p. 393).

As Reid saw it, Locke accepted both 1 and 2 and was therefore driven to deny 3, despite his affirmation of it elsewhere. He posited “abstract general ideas,” such as the infamous image of a triangle that is neither isosceles nor scalene, as merely generic entities existing in the mind. Berkeley and Hume, on the other hand, both accepted 2 and 3, and were thus led to reject 1. Not believing that entities such as Locke’s merely generic triangle could exist even in the mind, they denied that we are ever aware of general entities. Thus were born their attempts to explain how we can think generally (for example, in proving propositions about all triangles) by means of ideas that are particular.

Reid’s novelty is to deny proposition 2, which he castigates as one of the prejudices giving rise to the theory of ideas. It led all three of his predecessors in the British Empiricist tradition to affirm that the immediate object of awareness, in conception as well as in perception, must be an idea. By denying 2, Reid was enabled to uphold both 1 and 3, thus collecting together the truths his predecessors had divided between them.

## FIRST PRINCIPLES

Essay VI of the *Intellectual Powers* contains an extensive and important treatment of what Reid calls first principles. A first principle is a self-evident proposition—a proposition that is evident to us without need of any rea-

sons to support it. Like Aristotle, Reid thinks that our knowledge must ultimately rest on first principles, for without them we would be faced with an infinite regress of supporting propositions. He may therefore be classified as an epistemological foundationalist. Reid believes there are first principles both of necessary and of contingent truths. The first principles of necessary truths include axioms of logic, mathematics, grammar, metaphysics, and morals. The first principles of contingent truths include principles pertaining to the deliverances of consciousness (Reid's term for introspection), perception, memory, inductive reasoning, and others of our faculties.

In Reid's enumeration of the first principles of contingent truths, there is a subtle ambiguity that greatly affects how his epistemology is to be interpreted. Here is how he formulates Principle 1, which gives us the first principle(s) regarding consciousness: "First, then, I hold, as a first principle, the existence of every thing of which I am conscious" (EIP 6.5, p. 470). Putting this in terms of truth rather than existence, he might just as well have said that he holds, as a first principle, the truth of every proposition to which consciousness testifies. The ambiguity in Principle 1 may then be brought out by the following two ways of symbolizing it, where "Cp" is short for "I am conscious that p":

1.1 It is a first principle that  $(p)(Cp \rightarrow p)$ . (It is a first principle that for any proposition p, if I am conscious that p, then p.)

1.2  $(p)(Cp \rightarrow \text{it is a first principle that } p)$ . (For any proposition p, if I am conscious that p, then it is a first principle that p.)

The difference between the formulations is this: 1.1 says that it is a first principle that all the deliverances of consciousness are true. In other words, 1.1 gives us *one general proposition* as first principle. 1.2, on the other hand, says that each of the deliverances of consciousness—which may include propositions such as *I am now in pain*—are themselves first principles. So 1.2 gives us *many particular propositions* as first principles. A similar ambiguity holds in regard to the first principles of perception and memory.

How should Reid's first principles be understood—as general or particular? Perhaps the best overall interpretation of Reid's epistemology is provided by the particularist construction. At the very least, his epistemology must be understood as recognizing particular first principles even it recognizes general first principles as well.

If Reid's first principles are construed in the particularist way, he is not only a foundationalist in his epistemology but also (in one important sense) an *externalist*. Externalists hold that there are sources or factors that give a subject knowledge even if the subject does not know anything about how the factors work or whether they are reliable. On the particularist construction of Reid's principles, the mere fact that a proposition is a deliverance of perception, memory, or consciousness suffices to make that proposition evident (and thus, in favorable circumstances, known). To know that there is a tree over there, for example, one need only have a perception of a tree. It is not necessary for the subject to know anything about the reliability of sense perception, which may be a matter to which he has never given any thought. On the generalist construction of the principles, by contrast, the subject would presumably have to know that the general principles are true in order for knowledge of particular propositions to arise in accordance with them. In other words, he would have to know his faculties are reliable before they could be sources of any other knowledge. That puts an obstacle in the way of knowledge that skeptics might claim to be insurmountable.

It may be useful to summarize by drawing together the various things Reid has to say in response to skepticism about the material world. First, what skeptical philosophers profess cannot be believed and is not believed even by skeptics themselves. This is a point that Hume famously admitted, and it may be questioned what force it has against the truth or reasonableness of the skeptic's position. Second, the argument that we cannot even conceive of a material world is answered by Reid's nativism, according to which we are endowed by our constitution with conceptions of extended external objects.

Third, the argument that knowledge of the external world must be based on problematic inferences from our ideas is undercut by Reid's attack on the theory of the ideas. Fourth, the position of the "semiskeptical," who says we can be certain about the deliverances of our consciousness but not about anything else, is objectionably arbitrary. For who can prove that consciousness never errs? And what reason is there to believe the deliverances of consciousness that is not a reason for believing the deliverances of our other faculties as well? (Inq. 5.7, p. 71).

Finally, the position of the total skeptic, who refuses to accept the deliverance of any faculty unless its reliability is proven in advance, is irrefutable (Inq. 5.7, p. 71; cf. IP 6.5, p. 480). We cannot *show* that he or she is wrong without assuming something he or she would question.

But perhaps, for all that, we may *know* that he or she is wrong, if Reid's externalist approach to epistemology is correct.

### CAUSATION AND FREEDOM

As noted above, Reid thinks we have many conceptions, such as that of a self or subject of mental operations, that we could not have on Humean principles. Among them is the conception of active power, or real efficacy in bringing about changes, to which Reid turns in his third book. He thinks we obtain a clear conception of such power when we are conscious of our own activity in bringing something about by an act of will. Active power is exercised only by agents or substances, not by events, so in the strictest sense of causation, only agents are causes for Reid. When we speak of one event causing another, Reid tells us, it would be more proper to speak of events related by lawful sequence or a relation of sign and thing signified.

That we sometimes act freely (or that we possess "moral liberty") is, according to Reid, a natural conviction, comparable to our belief in a material world. In the *Active Powers*, he offers three arguments that we really do possess such liberty. The first is based on the "naturalness" of our conviction in regard to it, the second on the notion of moral responsibility, and the third on our ability to secure ends by prosecuting a long series of means.

Reid rejects accounts of moral liberty such as those of Hobbes and Hume, who seek to make liberty compatible with determinism. He would reject Hume's suggestion that I did A freely if I did A as a result of willing to do it and would have done otherwise if I had willed otherwise. In a universe in which my willing was itself the end of a causal chain stretching back to the Big Bang, the conditions of this definition might be satisfied, yet I would not, according to Reid, have acted freely. It is a further requirement of liberty that my willing not have been determined by antecedent events in that way. But that is not to say that my willing must be random or uncaused. On the contrary, in a case of free action, it is caused by *me*, the agent. In this way Reid brings his theory of agent causation into his account of liberty, attempting to escape the dilemma that has determinism as one horn and arbitrary uncaused acts of will as the other. Reid believes that every event has a cause, but he holds that the cause of an event need not be another event—it may be an agent.

Agent-causation theories of human action inspired by Reid began to undergo a revival during the last third of the twentieth century, finding advocates in Roderick Chisholm and Richard Taylor, among others. Such theo-

ries offer a tantalizing glimmer of hope for resolving old problems yet face formidable problems of their own. If I am the cause of my willing to do A, mustn't there be such an event as my causing the willing? If so, what is the cause of *that* event? If it is *nothing*, we have fallen back on the randomness horn and violated Reid's professed belief that every event has a cause. If it is a *further event*, we are back on the horn of determinism. If it is the *agent*, we have taken the first step of an infinite regress in which I am the cause of my willing A, the cause of my causing of my willing A, and so on, *ad infinitum*.

### MORAL PHILOSOPHY

Reid is often considered to be a member of the moral-sense school of philosophy, insofar as he holds that moral notions and moral determinations are the product of a moral faculty or sense. He insists, however, that the employment of the term *sense* is accurate only with the proviso that a sense can deliver judgments as well as feelings. In opposition to Hume, he holds that "moral approbation implies a real judgment" (AP 5.7, pp. 457–481), capable of being true or false, and is not merely an expression of feeling like "Hurrah!" (It must be said, however, that his criticisms of Hume sometimes convert the supposedly noncognitivist view he is attacking into a subjectivist form of cognitivism). In further opposition to Hume, he holds that reason is not merely the slave of the passions but has a real role to play in the selection of ultimate ends of action (AP, 5.3).

Reid also opposes another kind of view that sometimes goes under the rubric of moral-sense theory: the view that moral properties are analogous to secondary qualities, as in the suggestion that for an action to be right is for it to arouse favorable moral emotion in those who contemplate it. Reid protests that such accounts abolish the necessity of moral principles. It is necessary, according to him, that actions of certain types are right but contingent that they produce whatever effects they do in those who contemplate them (IP, 6.6, pp. 494–495). On the whole, Reid's views probably bear less resemblance to moral-sense theories than they do to the intuitionism of G. E. Moore. Much of what Reid says about *right* anticipates what Moore said about *good*: that it is indefinable, that we understand what it is by an original power of the mind, and that our moral faculty provides us with first principles about which types of acts are right and which wrong.

*See also* Aristotle; Arnauld, Antoine; Berkeley, George; Causation; Chisholm, Roderick; Common Sense;

Descartes, René; Ducasse, Curt John; Geometry; Hobbes, Thomas; Hume, David; Introspection; Locke, John; Malebranche, Nicolas; Meinong, Alexius; Moore, George Edward; Nativism, Innatism; Russell, Bertrand Arthur William; Smith, Adam.

## Bibliography

### WORKS BY REID

- The Works of Thomas Reid, D. D.*, edited with notes by Sir William Hamilton. 2 vols. 8th ed. Edinburgh, U.K.: James Thin, 1895. Reprint Hildesheim: Georg Olms, 1983.
- Essays on the Active Powers of Man*, edited by Baruch Brody. Cambridge, MA: MIT Press, 1969.
- An Inquiry into the Human Mind on the Principles of Common Sense*, edited by Derek R. Brookes. Edinburgh, U.K.: Edinburgh University Press, 1996. Reprint University Park: Pennsylvania State University Press, 1997.
- Essays on the Intellectual Powers of Man*, edited by Derek R. Brookes. University Park: Pennsylvania State University Press, 2002.
- The Correspondence of Thomas Reid*, edited by Paul Wood. University Park: Pennsylvania State University Press, 2002.

### WORKS ABOUT REID

- Alston, William P. "Thomas Reid on Epistemic Principles." *History of Philosophy Quarterly* 2 (1985): 435–452.
- Barker, Stephen F., and Tom L. Beauchamp, eds. *Thomas Reid: Critical Interpretations*. Philadelphia: Temple University Philosophical Monographs, 1976.
- Broadie, Alexander. "Reid Making Sense of Moral Sense." *Reid Studies* 1 (1998): 5–16.
- Cummins, Phillip. "Reid's Realism." *Journal of the History of Philosophy* 12 (1974): 317–340.
- Dalgarno, Melvin, and Eric Matthews, eds. *The Philosophy of Thomas Reid*. Dordrecht, Netherlands: Kluwer, 1989.
- De Bary, Philip. *Thomas Reid and Skepticism: His Reliabilist Response*. London: Routledge, 2002.
- Daniels, Norman. *Thomas Reid's Inquiry: The Geometry of Visibles and the Case for Realism*. Palo Alto, CA: Stanford University Press, 1974.
- Falkenstein, Lorne. "Reid's Account of Localization." *Philosophy and Phenomenological Research* 61 (2000): 305–328.
- Gallie, Roger. *Thomas Reid and "The Way of Ideas"*. Dordrecht, Netherlands: Kluwer, 1989.
- Grave, S. A. *The Scottish Philosophy of Common Sense*. Oxford, U.K.: Clarendon Press, 1960.
- Haldane, John. "Thomas Reid: Life and Work." *American Catholic Philosophical Quarterly* 74 (2000): 317–344. This issue contains eleven articles devoted to Reid.
- Immerwahr, John. "The Development of Reid's Realism." *The Monist* 61 (1978): 245–256. This issue is entirely devoted to Reid and includes thirteen articles on his philosophy.
- Lehrer, Keith. *Thomas Reid*. London: Routledge, 1989.
- McKittrick, Jennifer. "Reid's Foundation for the Primary/Secondary Quality Distinction." *The Philosophical Quarterly* 52 (2002): 478–494. This issue includes a previously unpublished manuscript by Reid ("Of Power") and ten articles on his philosophy.

- O'Connor, Timothy. "Thomas Reid on Free Agency." *Journal of the History of Philosophy* 32 (1994): 605–622.
- Rowe, William L. *Thomas Reid on Freedom and Morality*. Ithaca, NY: Cornell University Press, 1991.
- Van Cleve, James. "Thomas Reid's Geometry of Visibles." *The Philosophical Review* 111 (2002): 373–416.
- Wolterstorff, Nicholas. *Thomas Reid and the Story of Epistemology*. Cambridge, U.K.: Cambridge University Press, 2001.
- Woudenberg, Rene van, and Terence Cuneo, eds. *The Cambridge Companion to Reid*. Cambridge, U.K.: Cambridge University Press, 2004.
- Yaffe, Gideon. *Manifest Activity: Thomas Reid's Theory of Action*. Oxford: Clarendon Press, 2004.
- For complete bibliographies of Reid's own works and secondary sources, see *Reid Studies* 1 (1998): 71–81; 2 (1998): 57–78; and 4 (2000): 93–96.
- The Center for Scottish Philosophy, based at the University of Aberdeen, maintains the following web site with information about events and publications relating to Reid: <http://www.abdn.ac.uk/cssp/>

James Van Cleve (2005)

## REIMARUS, HERMANN SAMUEL (1694–1768)

Hermann Samuel Reimarus, the German philosopher and theologian, was born in Hamburg and studied theology at Jena. After serving as a lecturer in Wittenberg and as director of a high school in Wismar, he became a teacher of oriental languages at the Johannes-gymnasium in Hamburg. He began writing very late in life, when he was about sixty. One of his most important works, *Apologie oder Schutzschrift für die vernünftigen Verehrer Gottes* (Apology for or Defense of the Rational Worshiper of God), was first published by Gotthold Ephraim Lessing—posthumously and only in part—as fragments of an allegedly anonymous manuscript found in the Wolfenbüttel Library, where Lessing was librarian ("Wolfenbüttler Fragmente eines Ungenannten," in *Beiträge zur Geschichte und Literatur, 1774–1777*).

Reimarus was originally a Wolffian, and Wolffianism was a lasting foundation for his thought; but he developed individual doctrines in both philosophy and theology as one of the "popular philosophers." He stressed the moral aim of philosophy, that is, the happiness and moral perfectibility of man. He dissented from Christian Wolff chiefly in his views of philosophical methodology. He wrote in a "popular," or nonscholastic, style; he asserted that philosophy can be neither as certain as mathematics nor mathematically demonstrated; he stressed the func-

tion of common sense in knowledge; and he tried to simplify logic. In metaphysics, his main points of divergence from Wolff were his admission of a real interaction of soul and body and his view that life cannot be mechanically explained, but that it is an effect of the soul.

Reimarus's most important work was in the field of animal psychology and in his classification of the instincts of animals. Humans, unlike animals, have only a very few instincts. This lack may be a disadvantage for material life, but it is the basis for morality.

Reimarus appeared in his lifetime to be a moderate advocate of natural religion who did not openly oppose Christian revelation. But in the posthumous *Apologie* he submitted Christian revelation to a radical criticism in the spirit of English deism. In this work, for the first time in Germany the traditional view of Christianity was attacked neither on a speculative plan nor through superficial historical arguments, but on the basis of sound historical scholarship. Reimarus pointed out discordances between the Old and the New Testaments and between the different sections of each. He refused to accept the Gospels as the word of God, but described them as being an exposition of theological views elaborated by Jesus' successors in the leadership of Christianity. He considered the accounts of miracles, and in particular the account of the resurrection of Jesus, to relate events that never happened and to be forgeries of the Apostles. This purely rationalistic criticism made a tremendous impression on late eighteenth-century Germany, and deeply influenced the subsequent evolution of German theology.

**See also** Animal Mind; Deism; Lessing, Gotthold Ephraim; Wolff, Christian.

### Bibliography

#### ADDITIONAL WORKS BY REIMARUS

*Abhandlungen von den vornehmsten Wahrheiten der natürlichen Religion.* Hamburg, 1754.

*Vernunftlehre.* Hamburg and Kiel, 1756.

*Allgemeine Betrachtungen über die Triebe der Tiere.* Hamburg, 1760.

*Uebrigen noch ungedruckte Werke des Wolfenbüttelschen Fragmentisten*, edited by C. A. E. Schmidt. Berlin, 1786.  
Contains further portions of the *Apologie*.

#### WORKS ON REIMARUS

Aner, Karl. *Die Theologie der Lessingzeit.* Halle: Niemeyer, 1929; reprinted Hildesheim, 1964.

Büttner, W. H. *S. Reimarus als Metaphysiker.* Würzburg, 1909.

Engert, Joseph. *Der Deismus und die Religions- und Offenbarungskritik des H. S. Reimarus.* Vienna, 1915.

Engert, Joseph. *H. S. Reimarus als Metaphysiker.* Paderborn, 1909.

Fittbogen, G. "Lessings Entwicklung bis zur Begegnung mit Reimarus Werk." *Preussischer Jahrbücher* 73 (1918).

Richardt, H. *Darstellung der moralphilosophischen Anschauungen H. S. Reimarus.* Leipzig, 1906.

Schettler, R. *Die Stellung des Philosophen H. S. Reimarus zur Religion.* Leipzig, 1904.

Strauss, D. F. H. *S. Reimarus und seine Schutzschrift für die vernünftigen Verehrer Gottes.* Leipzig, 1862.

Giorgio Tonelli (1967)

## REINCARNATION

The doctrine variously called transmigration of souls, metempsychosis, palingenesis, rebirth, and "reincarnation" has been and continues to be widely believed. Although some of these terms imply belief in an immortal soul that transmigrates or reincarnates, Buddhism, while teaching rebirth, denies the eternity of the soul. The word *rebirth* is therefore the most comprehensive for referring to this range of beliefs.

In one form or another the doctrine of rebirth has been held in various cultures. It was expressed in ancient Greece (Pythagoras, Empedocles, Orphism, Plato, and later, Plotinus); among some Gnostics and in some Christian heresies such as the medieval Cathari; in some phases of Jewish Kabbalism; in some cultures of tropical Africa; and most notably in such Eastern religions as Jainism, Buddhism, Hinduism, and Sikhism. Some European philosophers, notably Arthur Schopenhauer and J. M. E. McTaggart, have incorporated the doctrine into their metaphysics. The origin of the doctrine of rebirth as a religious belief is obscure. There is evidence, both in Greece and India, that it was not characteristic of early Aryan cultures. It is virtually certain that in India it goes back to prehistoric times; it was then taken up by Brahmanic religion and appears as a new doctrine in the Upaniṣads.

Views vary about the scope and mechanism of rebirth. It is part of Indian thought, for instance—but not of African beliefs—that men can be reborn as animals and even as plants (not to mention as gods and spirits). Rebirth can take place not merely on Earth but also in a multiplicity of heavens and purgatories. Thus, although the prevalent belief is that rebirth occurs immediately upon death, this does not entail immediate earthly reincarnation, a feature that helps to make rebirth theory incapable of empirical disproof. In the Buddhist *Tibetan Book of the Dead*, however, a transitional period (*bardo*) of forty-nine days between death and rebirth is postulated. During this state the individual is translated to a

realm where he perceives the divine secrets; for the impure, these are so frightening that they flee back to earth and are reborn.

In Indian thought, there is a fairly large amount of speculation about the embryological mechanics of rebirth. Thus the Sāṃkhya school of Indian philosophy holds that the mental aspect of a person bears the impression of previous deeds (karma) and that it accordingly becomes associated with a particular fetus. But since during the period of fetal development the growing body is not capable of supporting the mental aspect, a “subtle” (unobservably refined) body is postulated. Thus the continuous element throughout rebirth and until liberation is the mental aspect associated with the subtle body.

In Buddhism it is held that the fetus results from the interaction of the sperm and material in the mother. These combine in a suitable way when associated with conscious states, as a further element in the process, to produce the right sort of individual to fit previous karma. Broadly speaking, then, rebirth theory implies that the genetic endowment of a person does not fully determine his early development but that a mental or spiritual factor associates itself with a suitable organism at conception. Thus karma is often taken to function through the homing of a soul upon a morally and physically appropriate fetus. McTaggart, in urging this, uses the analogy of chemical affinities.

A number of arguments in favor of the theory have been propounded; they can be classified as metaphysical, empirical, and theological. It is convenient to record here those arguments that do not depend too closely on metaphysical conclusions peculiar to particular philosophers, such as the argument for rebirth as accounting for knowledge of the Forms, as in Plato, and the complex metaphysical argument in McTaggart that depends in part on his theory of causation.

In Indian sources, two main metaphysical arguments have been employed. It may be noted that there has been relatively little explicit discussion of the issue in Indian philosophy, since no school was concerned with denying the doctrine, except the Materialist school, which was extinct by medieval times. (1) A Buddhist argument can be expressed as follows. All states have prior causes; some conscious states are not caused by bodily states; therefore the first physically uncaused state of an individual must have a prior nonphysical cause. But the existence of God is not admitted; hence there must be an empirical conscious state prior to conception and birth. This argument applies indefinitely in a backward direction through previous births. (It may be noted that the argument is con-

sistent with the Buddhist denial of an eternal soul, since the mental states of an organism are no more permanent than the physical ones.) (2) There is a Hindu argument from the eternity of the soul, which has been used in modern times by Radhakrishnan. Souls are eternal, but the normal condition for a soul is to be associated with a body. Hence it is likely that the soul in the past and future has a virtually everlasting succession of bodies. Thus metaphysical arguments attempting to establish the eternity of the soul have been taken to imply preexistence as well as postexistence.

Empirical arguments are as follows. (3) Children have instinctive capacities, which suggests that there must be learning prior to birth. Similarly, it is sometimes argued that child geniuses, such as Wolfgang Amadeus Mozart, indicate prenatal training. (4) Some people claim to remember past births, as in the case of Bridey Murphy. This claim is commonly made in the East for yogis and persons of deep spiritual insight, such as the Buddha and Buddhist saints. (5) The *déjà vu* experience and claims to knowledge of people and places that are not based on previous experience in this life have been cited as indicating rebirth. A counterargument is used against the objection that most people have no memories of such previous lives: Death is a traumatic experience (and so is birth), likely to cause amnesia. (6) The soul is indivisible and thus cannot derive from the parents, since it would then have to be a combination of parts.

The three important forms of theological argument are as follows. (7) Hindu and other scriptures and theologians are reliable in other matters and so ought to be reliable with respect to the teaching of rebirth. (8) Rebirth, associated with karma, provides a solution to part of the problem of evil, since inequalities and sufferings are the result of people’s past deeds. (9) The doctrine of rebirth provides the possibility of a long process of self-perfection, which harmonizes well with the religious vision of the world as a theater for moral striving.

The following are the objections that have been or can be brought against the arguments for reincarnation. Three objections to argument (1) are, first, the concept of emergent characteristics obviates the difficulty in explaining the cause of psychical states, although perhaps at the expense of being obscure. Second, the first premise (that all states have prior causes) is arguable, and it might be that nonphysically caused mental states are simply not caused. Third, the existence of God cannot be ruled out. (2) The plausibility of the argument depends on the plausibility of arguments for the eternity of the soul. Further, in Indian religious thought there is the possibility of



*mokṣa*, or *nirvāṇa*, a state of liberation in which there is no more rebirth. Consequently, it is inconsistent to hold that embodiment is necessary to souls. The Buddhist denial of a permanent self occasioned the criticism that there is nothing carried over to another life that would ensure individual continuity—the reply being that, on the Buddhist analysis, the individual in his present life is only a series of events, so that there is no essential difference in considering a succession of lives as constituting an individual series.

The following are objections to the empirical arguments. (3) Modern biology can sketch alternative explanations of instinct and genius in children. (4) Although some people seem to remember past lives, the evidence is not so unambiguous as to be conclusive; and if saintliness is a condition for remembering previous births, it would be difficult to verify such a memory—it would be hard to conduct an “experiment” in becoming a saint. (5) Similar problems arise with the evidence of *déjà vu* experiences. As to whether death is a traumatic experience, there is no evidence. (6) The creation of souls by God is compatible with the argument concerning the indivisibility of the soul; but in any case the argument depends on a soul-body distinction that may not be acceptable.

The objections to theological arguments are the following. (7) The validity of particular scriptures and theologies on matters of detail is especially suspect. (8) The argument that rebirth explains the existence of evil could not by itself be conclusive, since the problem of evil exists only for those who believe in a good God. (9) A similar consideration applies to the argument that rebirth allows the possibility of self-perfection.

Although believers in rebirth have scarcely touched on the matter, the theory of evolution also presents considerable difficulties to the traditional doctrine of a virtually infinite series stretching back into the past. In Indian mythological cosmology, however, there are periodic destructions of the cosmos, and during these periods embodied souls continue to exist latently; no doubt a similar assumption may deal with the above biological difficulties by arguing that before the emergence of life, souls existed latently, or in other parts of the cosmos. The problem remains, however, that this account would not be easily, if at all, checked by empirical evidence.

The hypothesis of reincarnation presents interesting problems about personal identity. If personal identity is analyzed in terms of memory, there would seem to be only a vacuous distinction between saying that *A* is reborn as *B* and that *A* and *B* are separate persons. C. J. Ducasse, however, has argued (*A Critical Examination of*

*the Belief in a Life after Death*, p. 225) that memory of any given life may be regained at some time or other in the series, and this would hold the series together. If bodily identity were held to be necessary to personal identity, rebirth could scarcely be meaningful, as it involves causal action at a distance in the transition from *A*'s death to *B*'s birth or conception.

**See also** Buddhism; Ducasse, Curt John; Empedocles; Evil, The Problem of; Gnosticism; Immortality; Indian Philosophy; Kabbalah; Karma; McTaggart, John McTaggart Ellis; Nirvāṇa; Orphism; Plato; Plotinus; Pythagoras and Pythagoreanism; Schopenhauer, Arthur.

### Bibliography

- Brandon, S. F. G. *Man and His Destiny in the World's Religions*. Manchester, U.K., 1962.
- Dasgupta, S. N. *History of Indian Philosophy*, 5 vols. Cambridge, U.K.: Cambridge University Press, 1922–1955.
- Davis, Stephen, ed. *Death and Afterlife*. New York: St. Martin's Press, 1989.
- Ducasse, C. J. *A Critical Examination of the Belief in a Life after Death*. Springfield, IL: Charles C. Thomas, 1961.
- Edwards, Paul. *Reincarnation: A Critical Examination*. Amherst, MA: Prometheus, 1996.
- Hick, John. *Death and Eternal Life*. New York: Harper and Row, 1976.
- Kear, Lynn. *Reincarnation: A Selected Annotated Bibliography*. Westport, CT: Greenwood Press, 1996.
- McTaggart, J. M. E. *Some Dogmas of Religion*. London: Arnold, 1906.
- O'Flaherty, Wendy, ed. *Karma and Rebirth in Classical Indian Traditions*. Berkeley: University of California Press, 1980.
- Radhakrishnan, S. *The Brahma Sūtra*. London, 1959. Translation with commentary plus a lengthy introduction to Vedānta.
- Reichenbach, Bruce. *The Law of Karma*. London: Macmillan, 1990.
- Smart, N. *Doctrine and Argument in Indian Philosophy*. London: Allen and Unwin, 1964. Ch. 12.
- Stevenson, Ian. *The Evidence for Survival from Claimed Memories of Former Incarnations*. New York, 1961.

**Ninian Smart (1967)**

*Bibliography updated by Christian B. Miller (2005)*

## REINHOLD, KARL LEONHARD (1758–1823)

Karl Leonhard Reinhold, the Austrian philosopher, was educated by Jesuits until the dissolution of their order in 1773, when he entered the Catholic college of the Barn-

abites, where he also taught, from 1778 to 1783. In 1783 Reinhold left Vienna for Leipzig and in the same year abandoned Catholicism in favor of Protestantism. A year later he moved to Weimar, where he was invited by Christoph Martin Wieland to contribute to his *Teutscher Merkur*. Soon he was not only Wieland's closest friend but also his son-in-law. Reinhold's first article, "Gedanken über Aufklärung," in which he traced the emergence of Enlightenment thought, appeared in July 1784, just a few months before the publication of Immanuel Kant's famous essay "What Is Enlightenment?" In his article Reinhold pleaded for the fuller realization of such. Enlightenment aims at greater tolerance toward religious minorities, more widespread secularization of knowledge and its greater accessibility to all sections of the population, and, above all, for the right of the individual to seek and assert truth free from fear, according to his critical reason and moral convictions.

Although two years later (1786) he was to publish a series of articles in support of Kant's critical philosophy, his second article in the *Merkur* (1785) was directed against Kant's unfavorable review of Johann Gottfried Herder's *Ideen*. The article appeared anonymously, but Reinhold later admitted his authorship to Kant. The articles dealing with Kant's *Critique of Pure Reason*, published under the title "Briefe über die Kantische Philosophie" from 1786 to 1787, established Reinhold's reputation as the most skillful exponent of Kant's philosophy and resulted in his being offered the chair of philosophy at the University of Jena in 1787. Reinhold was no less successful as a university teacher, and soon after his arrival Jena became one of the chief centers of Kantian studies. He attracted many students to Jena, and so great was his popularity that he was repeatedly urged to refuse the appointment offered him at the University of Kiel. Reinhold hesitated at first but eventually decided to move to Kiel in 1794, where he remained until his death.

One of the reasons for his departure, perhaps the most decisive, is revealed in a letter to Wieland that Reinhold later published in a selection of essays (*Auswahl vermischter Schriften*, Jena, 1796), under the title "Ueber die teutschen Beurtheilungen der französischen Revolution." Reinhold became increasingly worried over his countrymen's reactions to the excesses of the French revolutionary tribunals. In Kiel, which was then under Danish rule, he hoped to find a calmer political climate. Without condoning the terror of the revolutionaries, he nevertheless deplored the inferences that were drawn from it by leading public figures in Germany. In particular he viewed with anxiety the introduction of repressive measures and

the tendency to regard the French Revolution as a conspiracy of the philosophers. The French revolutionaries, he argued, may have been mistaken in attempting to deduce political rules from abstract principles that were often inadequately understood, but they were correct in their assessment of the desperate plight of their compatriots. If inferences were to be drawn, these would not suggest that philosophy presented a danger to orderly government but rather that disorderly government encouraged men to invoke philosophy in a manner unwarranted by its inherent limitations. Practical considerations such as these, no less than more strictly theoretical ones, prompted Reinhold to inquire more closely into the nature and scope of philosophical speculation.

Most of the works that he wrote at Kiel advanced a "fundamental philosophy" concerned with the basic presuppositions of scientifically valid thought. As the basic axiom of his "fundamental philosophy" Reinhold postulated the principle of consciousness, which he formulated in this way: By virtue of consciousness the perceiving (*erkennende*) subject is capable of distinguishing himself as something distinct from, while at the same time related to, the object of his consciousness, which, however, is not the object itself but rather the idea or notion (*Vorstellung*) of it. The consciousness itself constitutes a basic and irreducible fact, capable of neither proof nor further definition. It can only verify itself by reflecting upon itself. Reinhold was anxious to demonstrate that every thought process involves both a priori and a posteriori elements. The relation of the *Vorstellung* to the external object embodies its a posteriori material content (*Stoff*), whereas the subjective activity involved (*Vorstellungsvermögen*) in shaping the material content into a clear *Vorstellung* constitutes its a priori form (*Form*).

Reinhold stipulated three interconnected stages in the operation of consciousness: sense perception (*Anschauung*), which he classified as a receptive activity, and cognitive understanding (*Verstand*) and reflective reasoning (*Vernunft*), both of which he described as spontaneous activities. The product of these combined activities is the *Vorstellung*, which, Reinhold warned, must not be confused with an "image" or an "impression," for both terms suggest mere receptivity. Nor must it be identified with a "representation" of the object, since there is no way of either proving the identity of the *Vorstellung* with the object or even of comparing its similarity to the object. It follows that the object as such, no less than the subject as such, remains not only unknowable (as Kant realized) but also inconceivable. Both subject and object, therefore, as things-in-themselves are

pure abstractions. They are the residue of a *Vorstellung*, the thing minus the notion or conception of it.

Without denying the existence of things-in-themselves, Reinhold refused to commit himself as to the nature of their existence. He explicitly stated that he was merely anxious to determine the possibility and the limitations of cognition, not to inquire into its psychological origins or into the ontological nature of the objects of cognition. His declared aim was to provide a descriptive account, a phenomenology, rather than a theory of cognition, together with an analysis of the terminology commonly employed in this field. In spite of, or perhaps because of, Reinhold's deliberate delimitation of his theoretical undertaking, his works provided suggestively fertile starting points for subsequent Kantian research from Johann Gottlieb Fichte to Arthur Schopenhauer.

**See also** A Priori and A Posteriori; Consciousness; Enlightenment; Fichte, Johann Gottlieb; Herder, Johann Gottfried; Kant, Immanuel; Schopenhauer, Arthur; Toleration.

### Bibliography

#### ADDITIONAL WORKS BY REINHOLD

- Versuch einer neuen Theorie des menschlichen Vorstellungsvermögens*. Prague and Jena, Germany, 1789.
- Beyträge zur Berichtigung bisheriger Missverständnisse der Philosophen*. 2 vols. Jena, Germany, 1790 and 1794.
- Ueber das Fundament des philosophischen Wissens*. Jena, Germany, 1791.
- Versuch einer Kritik der Logik aus dem Gesichtspunkt der Sprache*. Kiel, 1806.
- Grundlegung einer Synonymik für den allgemeinen Sprachgebrauch in den philosophischen Wissenschaften*. Kiel, 1812.
- Das menschliche Erkenntnisvermögen aus dem Gesichtspunkt der durch die Wortsprache vermittelten Zusammenhänge zwischen Sinnlichkeit und dem Denkvermögen*. Kiel, 1816.

#### WORKS ON REINHOLD

- Adam, Herbert. *Carl Leonhard Reinholds philosophischer Systemwechsel*. Heidelberg: Winter, 1930.
- Bondeli, Martin, and Wolfgang H. Schrader, eds. *Die Philosophie Karl Leonhard Reinholds*. New York: Rodopi, 2003. In English and German.
- Cassirer, Ernst. *Das Erkenntnisproblem*. Vol. III, 33–57. Berlin: Cassirer, 1923.
- Klemmt, Alfred. *Karl Leonhard Reinholds Elementarphilosophie*. Hamburg: Meiner, 1958.
- Kroner, Richard. *Von Kant bis Hegel*. Vol. I, 315–326. Tübingen, 1921.
- Reinhold, Ernst. *Karl Leonhard Reinholds Leben und literarisches Wirken*. Jena, Germany: Frommann, 1825.
- Roehr, Sabine. "Freedom and Autonomy in Schiller." *Journal of the History of Ideas* 64 (1) (2003): 119–134.

Selling, Magnus. *Karl Leonhard Reinholds Elementarphilosophie in ihrem philosophischen Zusammenhang*. Lund, 1938.

Solomon, Robert C. *Introducing the German Idealists: Mock Interviews with Kant, Hegel, Fichte, Schelling, Reinhold, Jacobi, Schlegel, and a Letter from Schopenhauer*. Indianapolis: Hackett, 1981.

Von Schonborn, Alexander. "Karl Leonhard Reinhold: Endeavoring to Keep up the Pace Mit Unserem Zeitalter." In *The Emergence of German Idealism: Studies in Philosophy and the History of Philosophy*. Vol. 34, edited by Michael Baur. Washington, DC: Catholic University American Press, 1999.

**Frederick M. Barnard (1967)**  
*Bibliography updated by Tamra Frei (2005)*

## RELATIONS, INTERNAL AND EXTERNAL

Common sense would seem to hold that if some properties of a thing were taken away from it, it would no longer be the same thing. Further, it seems to hold that this is not the case for all properties of the thing. This intuition is the basis of the distinction between essential and accidental properties of a thing. It is also the basis of the distinction between the internal and the external relations that that thing bears to other things. For if among the properties that are essential to a thing (for example, the state of Maine) are relational properties, properties whose characterization essentially involves reference to some other thing (for example, the property of being north of Boston), then we say that the relations in question (for example, the relation between Maine and Boston) are *internal* to that thing (Maine). If we think that the thing would be the same were it (for example) not north of Boston—as in the case of a railroad car traveling through Maine—then we say that the relation in question is merely *external* to that thing.

The most familiar sort of relations considered when the topic of internal relations is discussed are relations between two or more particulars. However, the same internal–external distinction may be drawn in the case of relations between universals and particulars and also in the case of relations between two or more universals. If one holds that for every property *P* that a particular *X* displays, there is a universal, *P*-hood, to which *X* stands in the relation of "exemplification," then *all* of *X*'s properties may be construed as relational properties. Some of these relations of exemplification may be regarded as internal to *X* and others as external. Again, one may say that a universal such as "manhood" stands in an internal relation to certain other universals (for example, "rationality") and

in an external relation to other universals (for example, “philosopherhood”). Here the internal relation in question will be entailment, in the sense of “entails” in which we say that a given property (“being a man”) entails another property (“being rational”). In what follows, however, we shall confine ourselves as far as possible to relations holding between particulars, both because the philosophical literature has focused on such relations and because the notions of “exemplification of universals” and of “relations of entailment holding between universals” are sufficiently obscure and controversial to require detailed supplementary discussions. (Also, we shall not always trouble to distinguish between discussion of internal *properties* and of internal *relations*, since whatever doctrine a philosopher holds about the former will apply, *mutatis mutandis*, to the latter).

Two extreme positions have been put forward by philosophers who regard the internal-external distinction as unclear or incoherent. The first is that all of a thing’s properties are essential to its being what it is (and, a fortiori, that all its relations are internal to it). This position is associated with idealism and monism, for reasons that will emerge as we proceed. It holds that the connections between each of a thing’s properties (including its relational properties) and all of its other properties are so close that the deprivation of a single property would force us to say that, in a nontrivial sense, the thing is no longer what it was.

The second extreme position holds that none of a thing’s properties are essential to it (and thus, a fortiori, that no relations are internal to it). This view is put forward by those who make a firm distinction between the thing itself and a description of it. These philosophers say that, although certain properties of the thing are such that a given description could no longer be correctly applied to it were these properties absent, the notion that “the thing would no longer be the same” if these properties were absent is either trivial or misleading. For, in the weakest sense of “same,” the absence of any of its properties would make the thing no longer the same. Any stronger sense will, however, equate “being the same thing” with “being such that a given description correctly applies to it.” But since for each thing there are an infinity of equally correct descriptions, and nothing in the thing itself determines which of these is the description, any specification of “essential properties” will be arbitrary.

Both positions hold that the traditional essence-accident distinction, which is drawn by common sense and was first formulated explicitly by Aristotle, must be aban-

doned. The second position holds that the notion of “essential property” must be seen as a purely conventional notion, without a ground in the nature of the thing itself. It therefore suggests that we replace the notion of a relation being internal to a thing with the notion of a given relational description of a thing (such as “being north of Boston”) being internal to (that is, a necessary condition of) another description of the thing (such as “being in Maine”). The first position holds that the notion of “essential property” suggests, wrongly, that there is such a thing as a nonessential property. But since omniscience would see the universe as a seamless web (and, perhaps, as one single individual thing—the Absolute), this suggestion is misleading. Granted, they may say to representatives of the second position, that our present notion of “essential property” is a merely conventional one, we should not be led to conclude that things have no intrinsic natures. They do have intrinsic natures, but these can be known only *sub specie aeternitatis*, as facets of the Absolute. The commonsense essence-accident distinction is natural and inevitable, given the imperfect state of our knowledge. For omniscience, however, this distinction would be pointless.

This brief sketch of the opposing positions suffices to suggest how intimately the issues about internal relations are bound up with a whole range of other philosophical problems—problems about the notions of substance, of essence, and of “bare particulars,” about “real” versus “nominal” definitions, about nominalism versus realism, about the way in which we refer to and identify particulars, and about the nature of necessary truth. It is perhaps not too much to say that a philosopher’s views on internal relations are themselves internally related to all his other philosophical views.

#### THE VIEW THAT ALL RELATIONS ARE INTERNAL

The view that all relations are internal, in the form in which it has been discussed in the twentieth century, originated in the writings of the absolute idealist school in England and America in the period 1890–1920. In various forms it was held by F. H. Bradley, Josiah Royce, Bernard Bosanquet, and many others. Its most recent sustained defense is found in the work of Brand Blanshard, a follower of Bradley, notably in *The Nature of Thought* (1939). It has obvious historical connections with the doctrines of the seventeenth-century rationalists, notably Gottfried Wilhelm Leibniz’s view that all truths are analytic and Benedict de Spinoza’s assimilation of causal relations to logical relations. Its most important historical

antecedent, however, is the philosophy of G. W. F. Hegel. Hegel's insistence that the world was rational through and through, because Reason (or "Spirit") alone was real, was the principal inspiration of the philosophers who adopted the view that all relations were internal. For, if some relations were external, then the universe would be "impenetrable" to reason, in the sense that there would be brute particular facts not deducible from universal truths even by God himself.

A. C. Ewing, in *Idealism* (1934), provides a comprehensive account of the various meanings given to the term *internal* by exponents of absolute idealism and a critical analysis of those arguments in favor of the doctrine that all relations are internal that depend upon an ambiguous use of "internal." As Ewing points out, the meanings given to "internal" ranged from a very weak sense, in which to say that the relation *R* which *X* bore to *Y* was internal to *X* meant merely that "*R* makes a real difference to *X*," to a very strong sense, in which it meant that "from a knowledge of *Y* and *R* we could infer with logical necessity that *X* possesses a certain determinate or relatively determinate characteristic other than the characteristic of standing in the relation in question." Because such ambiguities permeate the discussion of the topic in such writers as Bradley and Royce, we shall not attempt an exegesis of their arguments. Instead, we shall attempt a reconstruction of two particularly persuasive arguments that seem to represent at least part of the common core of the absolute idealists' defense of their position on this subject. The two arguments to be examined by no means exhaust the repertoire of arguments that have been deployed in favor of the view that all relations are internal, but they are the arguments on which criticism of this view has chiefly centered.

**ARGUMENT FROM THE NATURE OF SELF-IDENTITY.** The first argument, which will be called here the argument from the nature of self-identity, was first clearly formulated by a critic rather than a proponent of the view that all relations are internal. G. E. Moore, in a classic attack on this view ("External and Internal Relations"), suggests that "one thing which is always implied by the dogma that 'All relations are internal' is that, in the case of every relational property, it can always be truly asserted of any term *A* which has that property, that any term which had not had it would necessarily have been different from *A*." The argument in favor of this view is simply that, as Moore puts it, "if *A* has *P*, and *x* has not, it *does* follow that *x* is other than *A*." In other words, it is unquestionably true that

- (1) *A* has *P* entails that (*x* does not have *P* materially implies that *x* is other than *A*).

Contemplation of this truth, Moore suggested, led philosophers to say that "*A* could not be what it is (but would necessarily be something different) did it not have *P*."

Now, as Moore points out, the argument as it stands is fallacious. (1) does not permit the conclusion that

- (2) *A* has *P* materially implies that (*x* does not have *P* entails that *x* is other than *A*).

Only (2) would permit the conclusion that *A* would necessarily be a different particular did it not have *P*. The difference between (1) and (2) may be put by saying that all that (1) tells us is that *A* cannot both have and not have the property *P*, whereas (2) tells us that *A* could not be *A* unless it had *P*. (1) is trivial, whereas (2) blurs the commonsense contrast between essential and accidental properties (and thus between internal and external relations). As Moore puts it, "(1) asserts that if *A* has *P*, then any term which has not, *must* be other than *A*; (2) asserts that if *A* has *P*, then any term which had not, would necessarily be other than *A*." Moore notes that to confuse the two propositions, "you have only to confuse 'must' or 'is necessarily' with 'would necessarily be.'" This confusion, in turn, will lead one to confuse the (physically necessary but logically contingent) fact that *A* has *P* with a statement about what is *logically* necessary for something to be *A*. While not attempting to cite examples of this fallacy in the writings of the absolute idealists, Moore claimed that much of their willingness to adopt the view that all relations are internal was due to their having confused (1) and (2). Whether or not this fallacy played the role in their thought that Moore thought it did is less important, from a historical point of view, than the influence exercised by Moore's diagnosis. Philosophers in general tended to agree with Moore that the absolute idealists had been guilty of this confusion, and his essay was a turning point in discussion of the topic. Defenders of the thesis of the internality of all relations who came after Moore were forced to produce arguments against the main presupposition of Moore's argument—that the commonsense distinction between logically contingent propositions and logically necessary propositions was unobjectionable. Crudely put, one may say that before Moore's essay, defenders of the view that all relations were internal felt able to argue that simple reflection on commonsense criteria for self-identity led to the conclusion they desired. After Moore's essay, they were forced to attempt to undermine common sense by claiming that the distinctions

Moore had drawn were, though commonsensical, philosophically indefensible.

**ARGUMENT FROM THE NATURE OF CAUSALITY.** The above was the strategy adopted by Blanshard in his *The Nature of Thought*, in which he presents the second, far more important and profound, argument in favor of the doctrine that all relations are internal. This may be called the argument from the nature of causality. Moore, like most philosophers in the tradition of British empiricism, had taken for granted a distinction between physical necessity and logical necessity, a distinction between the sense in which it is necessary, given the laws of nature and the past history of the universe, that a given particle be located at a given point in space at a given time, and the sense in which it is not necessary, *simpliciter*. Traditional rationalism, on the other hand, had questioned this distinction. Although earlier absolute idealists had also rejected the distinction between two kinds of necessity, they had done so *en passant*. They had treated it as simply one more consequence of empiricism's uncritical acceptance of a commonsense metaphysics that, they claimed to have shown, was fundamentally incoherent. Blanshard, approaching the matter epistemologically rather than metaphysically, brought forward a battery of arguments designed to show that the acceptance of this distinction was the result of a mistaken Humean analysis of knowledge. By weakening this distinction and claiming that causal necessity (by virtue of which *A* had *P*) could not be separated from logical necessity (by virtue of which *A* was self-identical), he was able to argue that what Moore had viewed as a simple confusion was at worst a confused formulation of a vitally important insight.

In examining this second argument, it will again be convenient to look to its critics rather than to its defenders. Ernest Nagel, in a critique of Blanshard's *The Nature of Thought* titled "Sovereign Reason," restates and criticizes Blanshard's views on internal relations in a way that brings out very clearly their connection with Blanshard's treatment of causality. Blanshard, in turn, has replied to Nagel in the later chapters (particularly Ch. 12) of his *Reason and Analysis* (1963). A summary of the Blanshard-Nagel controversy will serve two purposes. It will trace the most recent line of defense adopted by defenders of the view that all relations are internal, and it will lead us to an understanding of why some philosophers claim that no relations are internal.

Blanshard puts forward, and Nagel quotes as a basis for criticism, the following version of the doctrine that all

relations are internal. Despite the ambiguities detected by Ewing, Blanshard holds that "the principal meaning" of this doctrine is clear and formulates it as follows:

- (1) that every term, i.e., every possible object of thought, is what it is in virtue of relations to what is other than itself;
- (2) that its nature is affected thus not by some of its relations only, but in differing degrees by all of them, no matter how external they may seem;
- (3) that in consequence of (2) and of the further obvious fact that everything is related in *some* way to everything else, no knowledge will reveal completely the nature of any term until it has exhausted that term's relations to everything else. (*Nature of Thought*, Vol. II, p. 452)

Nagel notes, and Blanshard would agree, that everything here turns on the notion of the "nature of a term." If the term's nature includes all its properties, then Blanshard is right. Nagel bases his general objections to Blanshard on the claim that this is a perverse use of "nature," since "it is quite clear that just what characters are included in an individual, and just where the boundaries of an individual are drawn, depend on decisions as to the use of language. These decisions, though motivated by considerations of practical utility, are *logically arbitrary*" (p. 275). Nagel, in other words, is saying that "the nature of *X*" consists of just those properties of *X* whose absence would cause us to cease using "*X*" to refer to *X* and that the selection of these properties is determined not by empirical study but by convention. The list of such properties is finite, whereas the list of the properties of *X* is potentially infinite. Nagel thus adopts what has become the standard empiricist view, first clearly formulated by A. J. Ayer in "Internal Relations," that to determine which properties of *X* are internal to it is merely a matter of determining which propositions about *X* are analytic and that determining *this* is simply a matter of consulting linguistic usage. To urge that the nature of a thing includes all its properties would, given this view, be to urge that *all* propositions about *X* are analytic. Both Nagel and Ayer treat this conclusion as a *reductio ad absurdum*.

In examining Blanshard's arguments, Nagel first takes up Blanshard's form of the argument from the nature of self-identity and disposes of it by drawing what is essentially Moore's distinction between the logically contingent fact that *A* has *P* and the logically necessary fact that anything that does not have *P* cannot be identical with *A*. His defense of this distinction is simply that unless the distinction is drawn, we shall wind up with the view that "the nature of *X*" is identical with *X* itself and

thus that “the nature of a thing, like the thing itself, would be something that is in principle indefinable and could not therefore be made the basis for bringing into systematic order any of the characters which the thing displays” (p. 276). But from Blanshard’s point of view, this reply begs the question, since Blanshard would be quite willing to say that the nature of any given particular is indeed indefinable (by finite minds). For Blanshard the question is merely pushed back to the issue of whether a satisfactory epistemology can be constructed on the basis of the view that all logical necessity has its source in linguistic convention. But this latter issue is just the issue of whether causal relationships (which are agreed on all sides to be matters not of convention but of empirical inquiry) can, in the last analysis, be held to be distinct from logical relationships. If they cannot, then it would seem fair to say that although we must (unfortunately) work with the commonsense distinctions between necessary and contingent truths, essence and accident, physical and logical necessity, and the like, these distinctions are nevertheless mere pragmatic makeshifts (pertaining, in Bradleian terminology, to Appearance rather than to Reality). To invoke them to is to attend not to how things are but merely to how we are forced (by the limitations of our minds and of our everyday language) to talk about them.

Thus the battle between Blanshard and Nagel is truly joined only when Nagel takes up the question whether “logical necessity is involved in causal relations.” Blanshard has, as Nagel notes, two principal arguments for the view that it is so involved. The first is that causal relations must be analyzed either in terms of “mere regularity of sequence” or in terms of “entailment.” The failure of the regularity view will, in Blanshard’s eyes, constitute a proof of the entailment view. But the entailment view is just that “*A* causes *B*” is a statement about a logical relation between *A* and *B*. Now if (as is not implausible) all true relational propositions about particulars are propositions that are true in virtue of causal relations between the particulars mentioned in these propositions, then it follows that all particulars are connected to all others by logical relations and that every such proposition would be seen (by omniscience) to entail a logical truth about every such particular.

Nagel has two objections to this argument. First, the “regularity” and “entailment” views do not exhaust the available analyses of causality; second, “the entailment view contributes *nothing* toward advancing the aims of specific inquiries into the causal dependencies of physical nature.” The second objection can be dismissed by Blan-

shard as irrelevant, since he is quite willing to admit, with David Hume, that observation of regular sequence is our only method for determining what causal relations actually hold (except, perhaps, in the case of “direct insight” into certain relations between mental states or events). Blanshard need merely insist that regularity provides evidence of an underlying entailment but that the regularity and the entailment must not be confused. Blanshard offers no reply to Nagel’s first objection, but one suspects that he would argue that all proposed *via media* analyses of causality in fact boil down to one of the two alternatives he has suggested. Even if this point is granted to Blanshard, however, the whole question of the validity of his attack on the regularity theory remains. We must leave the topic with the remark that Blanshard can, in attacking this theory, take full advantage of the embarrassment encountered by Rudolf Carnap, Nelson Goodman, and others in their attempts to construct an inductive logic on the basis of Neo-Humean “regularities.” Further, recent work in inductive logic (such as Goodman’s *Fact, Fiction and Forecast*, 1955) and the philosophy of science (the work of Hilary Putnam, Wilfrid Sellars, P. K. Feyerabend, and others) has made it apparent that the distinction between matters of convention and matters of fact is not so clear as Hume and the early positivists believed. This recent work is closely connected with W. V. Quine’s skepticism about the analytic-synthetic distinction and related work in the philosophy of language. It is perhaps not too much to say that empiricism is presently in a state of crisis and that the crisis revolves precisely around the validity of the distinctions that empiricists have traditionally invoked against the thesis of the internality of all relations. We must conclude that the question of the validity of Blanshard’s first form of the argument from the nature of causality must remain undecided until these issues have been further clarified.

Before leaving the Blanshard-Nagel controversy, however, we must take up the second of Blanshard’s arguments in favor of the view that logical necessity is involved in causation. This argument is that philosophical reflection upon the nature of causality leads us to conclude that

to say that *a* produces *x* in virtue of being *a* and yet that, given *a*, *x* might not follow, is inconsistent with the laws of identity and contradiction. Of course if *a* were a cluster of qualities abstracted from their relations, and its modes of causal behaviour were another set conjoined with the former externally, then one could deny the latter and retain the former with perfect con-

sistency. But we have seen that when we say *a* causes *x* we do *not* mean that sort of conjunction; we mean an intrinsic relation, i.e., a relation in which *a*'s behaviour is the outgrowth or expression of *a*'s nature. And to assert that *a*'s behaviour, so conceived, could be different while *a* was still the same would be to assert that something both did and did not issue from the nature of *a*. (*Nature of Thought*, Vol. II, p. 513)

With this argument, as Nagel notes, we are back at the perplexing notion of "the nature of *a*." Whereas the entailment analysis of the nature of causation can perhaps be stated without using the notion of the "nature of *A*" (although if it were, it might be difficult for Blanshard to infer the thesis of the internality of all relations from the truth of the entailment view), this present argument about the nature of causality makes essential use of this notion. At this point, therefore, Nagel returns to his general line of attack on Blanshard's formulation of the thesis of the internality of all relations and argues that what Blanshard says here is true only if "the nature of *X*" is defined as "all the properties of *X*," a definition that, in Nagel's eyes, is both idiosyncratic and such as to trivialize Blanshard's claim.

The effectiveness of Nagel's reply can be judged only in the light of a general theory about the relation between thought, language, and reality. For, here again, Nagel is taking for granted the view that whether a given property is included within a thing's nature is a question about our language, rather than a question to be settled by further inquiry about the thing itself. Just as judgment of the validity of the first form of Blanshard's argument from the nature of causality must be postponed until certain general philosophical issues have been (at least) clarified, so also judgment of the validity of the second form of this argument must be deferred until questions about the standard empiricist doctrine that all "essences" are "nominal" and that "real essence" is an incoherent notion are settled. For Blanshard can insist that Nagel has begged these latter questions. In *Reason and Analysis* we find Blanshard arguing that Nagel's view that decisions about what characters are included in an individual are "logically arbitrary" leads to the view that, for example, Socrates's snub-nosedness is as good a candidate for an essential property of Socrates as his philosopherhood. Blanshard thinks this a *reductio ad absurdum*, but this rebuttal, once again, merely moves the argument one step further back. Nagel's point is not that we arbitrarily select which characteristics of an individual shall count as essential but that the criteria of selection are pragmatic,

dictated by our present interests and the modes of classification that we have, in the past, found it convenient to adopt. Nagel would say that a choice about linguistic usage, which is, from a practical point of view, far from arbitrary, is nonetheless logically arbitrary, in the sense that a language with alternative conventions is, though inconvenient, perfectly possible.

Blanshard's basic disagreement with Nagel consists in his view that such pragmatic considerations are not the last word and his insistence that the goal of thought is the discovery of real essences. Such real essences would be discovered by discovering the chains of entailment that connect all the various universals that characterize (and, in Blanshard's metaphysics, constitute) a particular. In Blanshard's view, to say that analytic propositions are true by convention is thoroughly misleading, for such conventions are the results of attempts to discover such entailments. For Blanshard the identification of the nature of *X* with *X* itself, and of both with the totality of properties that characterize *X*, and of all of these with *X*-as-known-by-an-ideal-knower (one who could grasp the entailments between all of these properties), is not (as it is for Nagel) a series of confusions but is forced upon us by an analysis of what we mean by "knowing *X*." The validity of Blanshard's second form of the argument from the nature of causality ultimately depends upon the validity of this analysis.

UNIVERSALS. The nature and depth of the issues involved in the controversy between Blanshard and Nagel may be further clarified by calling attention to one more area of disagreement between them. This concerns the nature and knowledge of universals. Blanshard views a particular as a congeries of universals and views the internal relations between particulars as reflecting the internal relations holding between the universals that constitute them. It is almost a cliché of recent analytic philosophy that to have knowledge of a universal is simply to know the meaning of a word; thus, to be acquainted with all the universals that characterize a particular would be merely to know the meanings of all the words correctly applicable to that particular. Such knowledge would obviously fall far short of telling us about the relations in which that particular stands to other particulars. For Blanshard, however, universals have natures that are not known to those who merely know the meanings of the words that signify those universals. To know the nature of a universal "fully and as it really is" would involve knowing its relations to all the universals that are exemplified in all the particulars that exemplify the first universal.



Thus, to know any universal “fully and as it really is” would be possible only for omniscience, just as, and for the same reasons that, knowledge of the real essence of a particular would be possible only for omniscience. Thus, resolution of the controversy about internal relations would require, at a minimum, a decision concerning the adequacy of a nominalistic account of universals. Blanshard views the current antagonism toward idealism (and a fortiori toward the thesis of the internality of all relations) as largely a result of analytic philosophy’s “systematic confusion between thought and language,” a confusion that leads philosophers such as Ludwig Wittgenstein to hold (1) that the notion of having a concept or being acquainted with a universal prior to the use of language is incoherent, and (2) that the notion of detecting internal relations between universals apart from considerations of linguistic usage is a relic of a radically mistaken analysis of mental events. If these latter tenets are accepted, clearly Blanshard’s arguments cannot even get off the ground. Once again, we must conclude that the thesis of the internality of all relations cannot be profitably discussed until one has taken sides on the most fundamental issues in contemporary philosophy.

#### THE VIEW THAT NO RELATIONS ARE INTERNAL

When we turn to the view that no relations are internal, we turn from a controversy that reflects profound underlying disagreements concerning the analysis of knowledge to a controversy about much narrower issues concerning the analysis of naming and predication. Those who say that no particular is internally related to any other particular insist that the only entities that can be internally related to one another are characteristics of particulars. Following to its logical conclusion Nagel’s claim that the assignment of a given description to a given particular is “logically arbitrary,” they hold that to say that *X* would “not be what it is” unless it had *P* is merely to say that the particular could not be characterized in a given way unless it had this property. But since the particular is sublimely indifferent to how it is characterized, it “is what it is” regardless of whatever properties it may have. To speak of “logically necessary conditions for the self-identity of *X*” is, at best, to speak elliptically of “logically necessary conditions for correctly describing *X* as a *K*,” where “*K*” signifies some kind of thing of which *X* is a representative, or (more generally) of “logically necessary conditions for correctly describing *X* as *C*,” where “*C*” is some general characterization.

The whole notion of “properties (and, a fortiori, relations) such that *X* would cease to be what it is if they were removed” is thus either incoherent or misleading. For “being what it is” is simply too ambiguous a notion; there are indefinitely many kinds to which *X* belongs and indefinitely many characterizations that apply to it. “Being what it is” is incoherent if it suggests that one of these kinds or characterizations is *intrinsically* privileged and misleading if a user of the phrase has already picked out some such kind or characterization, thus making his choice “privileged” by stipulation. To philosophers who deny the internality of any relations, the whole notion of internal properties and relations is an unfortunate vestige of the Aristotelian notion that there are real essences of particulars to be discovered by empirical inquiry. These philosophers heartily agree with the seventeenth-century rationalists, and with Blanshard, that any Aristotelian attempt to divide intrinsically essential and intrinsically accidental properties is foolish. But whereas Blanshard, sticking to the quest for real essences, insists that this point merely shows that the real essence of an object must include *all* its properties, these philosophers take the point to show the incoherence of the notion of “real essence” and the notion of “internal property.”

It may be useful to put the contrast between the roughly Aristotelian commonsense view and the two extreme views in yet another way. If we say that common sense holds that there are both particulars and properties of particulars, then we may say that common sense holds that each particular stands in a necessary relation to some of its properties and in a contingent relation to others. Blanshard dissolves the particular into a congeries of properties, and, because he believes (a) that properties (qua universals) have intrinsic natures to be discovered by inquiry (other than inquiry into linguistic usage) and (b) that such inquiry would, in principle, discover relations of entailment between all possible properties of all possible particulars, he holds that a particular stands in a necessary relation to all its properties. Philosophers who deny both doctrines and who assert (c) that “logical necessity” can only characterize relationships between universals, naturally emerge with the conclusion that the whole notion of logically necessary relations between particulars and their properties must be discarded. To put it picturesquely, Blanshard thinks that the dissolution of the traditional essence-accident distinction leaves us with the particular as a node in a network of internal relations between universals. His opponents think that this dissolution leaves us with “bare” particulars on the one hand (particulars that could logically have any properties) and with a network of entailments between universals on the

other (a network that is, however, much “looser” than Blanshard’s, since between most universals no relations of entailment exist).

**CONCEPT OF BARE PARTICULARS.** As an illustration of the movement toward leaving particulars bare, we may cite Gilbert Ryle, who says, in his article “Internal Relations,” that

for this view [the thesis of the internality of all relations] to be true *or false*, it would have to be significant to predicate a logically proper name or designation of a logically proper name or designation; and it would have to be significant to assert or deny that *this* was *this*; and the question “is anything this?” would have to mean something.... “This” is not a predicate, and a sentence in which it pretends to function as one is meaningless. So there *could* be no such dispute as to whether this’s being this does or does not depend on its being in one or other of its relations. (p. 165)

This line of thought suggests the general conclusion that there are no analytic propositions that ascribe properties to particulars. For example, it is misleading to call “Socrates was a Greek philosopher” analytic, for what this statement expresses is either (1) the contingent fact that certain features (snub-nosedness, being married to Xanthippe, and so on) were compresent with certain others (being Greek, being a philosopher), or (2) the contingent fact that the word *Socrates* is used to refer to an individual who exhibited certain features.

Even among philosophers who both reject (*a*) and (*b*) and accept (*c*), however, this general conclusion has been a matter of debate. In what follows, we shall consider an attempt to avoid the conclusion that there can be no analytic propositions that ascribe properties to particulars and an attempt to avoid the extreme position that no relations are internal to particulars by providing a “rational reconstruction” of the commonsense view. Such attempts are motivated, at least in part, by philosophical discomfort over the notion of “bare particulars.” The nature of this discomfort may be illustrated by considering the question “What, then, *are* these particulars, apart from the properties we ascribe to them?” If particulars really are “bare,” then any answer to this question is bound to be either wrong (if it lists some features that are criteria for particularity) or unhelpful (if it consists in saying simply “Well, particulars are just the kind of thing that properties can be ascribed to”). Although the realistic bent of contemporary analytic philosophy makes

philosophers hesitate to accept the Bradleian-Blanshardian view that the whole category of (plural) “particulars” belongs to Appearance rather than to Reality, it nevertheless seems that having only bare particulars would be as bad as having no particulars at all.

**INTERNAL PROPERTIES AS RELATIVE.** The most explicit and comprehensive attempt to avoid Ryle’s conclusion and still retain most of his premises is found in an article by Timothy Sprigge (“Internal and External Properties”); an examination of Sprigge’s treatment of the problem will bring out the underlying issues concerning naming and predication upon whose resolution the present question depends. Sprigge notes that the strength of the Rylean position lies in the fact that

in sentences expressing particular propositions where the subject word is a name, the subject word has no connotation. Therefore no predicate word can have a connotation which is incompatible with the connotation of the subject word. But a subject-predicate sentence could only express a necessary proposition if the connotation of the subject word were incompatible with the connotation of the negation of the predicate word. ... Of course, this rests upon the questionable view that there may be naming words without connotation—and this indeed is basically the point at issue. (p. 204)

One reason why this latter point is disputable is, as Sprigge says, that “it seems that one must identify a thing by some description. Having been thus identified,” he continues, “as answering to that description, is it not in effect defined as the thing having those properties, which properties therefore it necessarily has?” (p. 205). In other words, proper names could not be used unless their users could identify their referents, and how could the users do this save by having a description in mind? Must we not say that the notion that the logician’s dogma that “proper names do not connote” is true only of such Russellian “logically proper names” as “this” (which cannot be used save in the presence of their referents)? Sprigge replies to this point by granting it but noting also that since the same particular can be identified by an indefinitely wide range of different descriptions, the point is useless if one is trying to defend the notion of internal properties. In the case of a predicate, rough agreement on criteria for its application is required if the term is to play a useful role. But there seems nothing to prevent every speaker of the language from having a different set of procedures for identifying a particular while nevertheless using the same

proper name for it. Too many connotations are, so to speak, as bad as no connotation at all for purposes of formulating necessary truths.

If we follow Sprigge here, we need not be troubled by the spectacle of bare particulars. Every particular we refer to will always be dressed in some description or other, so we need not worry about how they look when undressed. But since each particular can be dressed up in so many ways, we are as far as ever from understanding what an “internal property” might be, unless we relativize the notion and say that certain properties are internal to  $X$  relative to a person  $S$  whose personal criteria for identifying  $X$  include the presence of these properties. Relativizing the notion in this way is, in essence, the basis for Sprigge’s “reconstruction” of the notion of internal property. As a sample of the sort of intuition upon which the commonsense distinction between internal and external properties is based, he notes that even though we are driven by the Rylean reasoning outlined above to call all subject-predicate statements about particulars synthetic, we find it hard to imagine the falsehood of, for example, “Scott was, at some time in his life, a man.” But what is a synthetic proposition if not one whose falsehood can be imagined? Sprigge proposes that we simply face up to the fact that there is a class of propositions that, if we *must* choose between calling them synthetic or analytic, must be called synthetic, even though they do not have imaginable contradictories. Specifically, they are such that no program of empirical inquiry could be formulated that would lead us to decide between them and their contradictories. The point is most effectively made in the following passage:

To ask whether a thing could have been quite different, from what it is, whether Scott could actually have had all the properties of Handel, is on a different level. The questions we have just been asking are all to some degree requests for further descriptions of Scott. But the present question is not one that calls for any investigation of Scott, and it is difficult to accept that a question which calls for no investigation of Scott, to which nothing about Scott is relevant, is really about Scott. (p. 209)

On the basis of these considerations, Sprigge makes the following proposal:

I suggest that a property is internal to a particular to the extent that no information about that particular is conveyed by one who says that it might have lacked that property. I think that the distinction between internal and external prop-

erties is not exact.... Let  $F$  be any property of a thing  $a$ . Then  $F$  is an external property of  $a$  if something interesting and true may be said of the form “if such and such then not- $Fa$ .” Otherwise  $F$  is an internal property. But as from different points of view different things are interesting, so from different points of view different properties are internal and external. (p. 210)

The notion of “internal” is thus not only made a matter of degree but also relativized to the interests and purposes of those who are discussing  $X$ . Conceivably, everyone might be interested in  $X$  for a widely different reason; in this case, it would be quite possible that everyone might identify  $X$  by means of a widely different, but equally true, description. Then there would be no agreement on internal properties, and an Aristotelian metaphysics would seem unintelligible to us. As it stands, however, we tend to be interested in things for roughly the same reasons and thus to group the same things into the same natural kinds (for example, to regard Scott as “essentially” a man, rather than as a collection of physical particles occupying a given stretch of space time, or as a colorful patch on the landscape of nineteenth-century Scotland). Given this agreement and given our natural taxonomical instincts (our tendency to turn differences of degree into differences of kind whenever possible in order to facilitate inquiry), we can explain the commonsensical character of the distinctions between essence and accident and between internal and external properties (and, a fortiori, internal and external relations).

As an account of the internal-external distinction that avoids both the arbitrariness of Aristotelianism and the counterintuitive character of absolute idealism, Sprigge’s proposal is a happy solution. But, like all such solutions, it is no better and no more permanent than the conceptual framework within which it is constructed. There is, to put it mildly, no consensus among philosophers of language as to when a sentence is “about” a given particular, when two sentences are about the same particular, the proper analysis of the notion of “name,” the reducibility of names to descriptions, the assimilation of demonstrative pronouns to proper names, the question of whether proper names can be said to have meanings, the utility of the analytic-synthetic distinction, the equation of “necessary truth” with “analytic truth,” and a host of related issues. In the absence of a comprehensive philosophy of language in which these issues are clarified and resolved in a systematic way, Sprigge’s proposal must be treated as a useful guideline, rather than as a definitive

resolution of the issue concerning internal relations. One can imagine, for example, a revivification of the Aristotelian doctrine of predication, according to which "Socrates is a man" exemplifies a radically different sort of predication from "Socrates is a Greek," such an Aristotelian philosophy of language would, when conjoined with a realistic, anti-instrumentalist philosophy of science, produce a view according to which it would make good sense to say that Socrates's humanity really was internal to him, not simply relative to our interests but absolutely and intrinsically. Such a view would argue that "man" signifies a natural kind and is thus naturally suited to be a predicate "in the category of substance," whereas "Greek" or "atoms located at  $p$  at  $t$ " is not, and that this is an empirical truth.

There probably would never have been a problem about internal relations were it not for the efforts of speculative metaphysicians, such as Parmenides, Spinoza, and Hegel, to undermine our commonsense conceptual framework. If one rejects such attempts out of hand, one will treat the adoption of monism and of the thesis of the internality of all relations as a *reductio ad absurdum* of the premises from which these views are derived. Since Moore, the vast majority of Anglo-American philosophers have rejected such attempts and have differed only in their diagnoses of the confusions of falsehoods that engendered metaphysical conclusions. As long as the dogma that logical necessity was a matter of linguistic convention remained unchallenged, a simple and elegant resolution of the problem of internal relations seemed possible. However, recent doubts about this dogma (combined with the realization that Aristotle's distinction between essential and accidental properties is not simply a philosopher's invention but is firmly grounded in common sense) have made the problem look more complex than it appeared in the days of Ayer's *Language, Truth and Logic*. Philosophers who wish, as P. F. Strawson has put it, to substitute a "descriptive" metaphysics for a "revisionary" one are now faced with the problem of reconciling (a) the existence of this commonsense distinction with (b) the standard empiricist view that knowledge of how we speak either does not reveal anything about the nature of the objects we refer to, or at least does so in a very different way than does empirical research directed to those objects themselves, (c) the fact that the meaning we assign to a term is in part a function of the amount of empirical knowledge we possess, and (d) the fact that common sense seems to require a realistic, rather than an instrumentalistic, view of what it is to "know the nature of an object."

If the difficulties of such a reconciliation prevent "descriptive" metaphysicians from carrying out their chosen task, then the door will be open once again to the two extreme views examined above. It may turn out that common sense is, if not as incoherent as Parmenides and Bradley thought it, at least sufficiently inconsistent as to require the adoption of paradoxical philosophical theses. Whether one then turns to the extreme represented by Ayer's radical conventionalism and instrumentalism, or to the extreme represented by Blanshard's idealistic monism, will be largely a matter of taste. Both views, as suggested above, are parts of internally consistent philosophical systems. Each system retains certain portions of our commonsense framework and insists on these at the expense of other portions. In the absence of a touchstone other than common sense, it is difficult to see how a rational choice between such systems can be made.

**See also** Absolute, The; Aristotle; Ayer, Alfred Jules; Blanshard, Brand; Bosanquet, Bernard; Bradley, Francis Herbert; Carnap, Rudolf; Common Sense; Goodman, Nelson; Hegel, Georg Wilhelm Friedrich; Leibniz, Gottfried Wilhelm; Moore, George Edward; Nagel, Ernest; Parmenides of Elea; Putnam, Hilary; Quine, Willard Van Orman; Realism; Royce, Josiah; Ryle, Gilbert; Sellars, Wilfrid; Spinoza, Benedict (Baruch) de; Strawson, Peter Frederick; Universals, A Historical Survey.

### Bibliography

- Allaire, Edwin B. "Bare Particulars." In *Essays in Ontology*. Iowa Publications in Philosophy, Vol. I, 14–21. The Hague, 1963.
- Alston, William P. "Internal Relatedness and Pluralism in Whitehead." *Review of Metaphysics* 5 (1951–1952): 535–558.
- Anscombe, G. E. M. "Aristotle." In *Three Philosophers*, by G. E. M. Anscombe and P. T. Geach. Ithaca, NY: Cornell University Press, 1961. Gives a sympathetic treatment of Aristotle's distinction between secondary substance and quality.
- Ayer, A. J. "Internal Relations." *PAS*, supp., 14 (1935): 173–185. Reprinted in *Ayer's Language, Truth and Logic*. London: Gollancz, 1936. Ch. 8.
- Blanshard, Brand. *The Nature of Thought*, 2 vols. London: Allen and Unwin, 1939; New York: Macmillan, 1940.
- Blanshard, Brand. *Reason and Analysis*. La Salle, IL: Open Court, and London: Allen and Unwin, 1962.
- California, University of. *Studies in the Problem of Relations*. University of California Publications in Philosophy, Vol. XIII. Berkeley, 1928.
- Chappell, Vere. "Sameness and Change." *Philosophical Review* 69 (1960): 351–362. On criteria of self-identity.
- Church, Ralph W. "On Dr. Ewing's Neglect of Bradley's Theory of Internal Relations." *Journal of Philosophy* 32 (1935): 264–273.

- Ewing, A. C. *Idealism*. London: Methuen, 1934. This contains the best exposition of the similarities and differences between the exponents of absolute idealism and also of the relation between the doctrine that all relations are internal and other idealistic doctrines.
- James, William. *Essays in Radical Empiricism*. New York: Longman, 1912. Ch. 3, "The Thing and Its Relations." On Bradley.
- Moore, G. E. "External and Internal Relations." *PAS* 20 (1919–1920): 40–62. Reprinted in Moore's *Philosophical Studies*. New York: Harcourt Brace, 1922.
- Nagel, Ernest. "Sovereign Reason," in his collection of articles *Sovereign Reason*. Glencoe, IL: Free Press, 1954.
- Rome, Sydney, and Beatrice Rome, eds. *Philosophical Interrogations*, 219–246. New York: Holt, Rinehart and Winston, 1964. Largely a continuation of the debate between Nagel and Blanshard.
- Royce, Josiah. *The Spirit of Modern Philosophy*. Boston: Houghton Mifflin, 1892. Expounds the relation between early rationalism and Hegel, as it was conceived by the absolute idealists.
- Russell, Bertrand. "The Monistic Theory of Truth." In his *Philosophical Essays*, 150–169. London: Allen and Unwin, 1910. Criticism of absolute idealism.
- Ryle, Gilbert. "Internal Relations." *PAS*, supp., 14 (1935): 154–172.
- Sellars, Wilfrid. *Science, Perception and Reality*. London, 1963. Ch. 9, "Particulars." On criteria for self-identity of particulars.
- Sprigge, Timothy, "Internal and External Properties." *Mind* 71 (1962): 197–212.
- Stace, W. T. *The Philosophy of Hegel*. London: Macmillan, 1924. Reissued New York: Dover, 1955. Ch. 2 discusses the relation between early rationalism and Hegel, as it was conceived by the absolute idealists.
- Thompson, Manley. "On the Distinction between Thing and Property." In *The Return to Reason*, edited by John Wild. Chicago: Regnery, 1953. Another defense of Aristotle's distinction between "secondary substance" and "quality."
- Will, Frederick. "Internal Relations and the Principle of Identity." *Philosophical Review* 44 (1940): 497–514.
- Wollheim, Richard. *F. H. Bradley*. Harmondsworth, U.K.: Penguin, 1959. Ch. 3 contains an exegesis of Bradley's treatment of relations.

**Richard M. Rorty (1967)**

## RELATIVISM IN ETHICS

See *Ethical Relativism*

## RELATIVITY OF KNOWLEDGE

See *Functionalism in Sociology; Historicism; Skepticism, History of; Sociology of Knowledge*

## RELATIVITY THEORY

From 1905 to 1915 Albert Einstein revolutionized the conception of space and time and gravity that had been central in physics since Isaac Newton. For a brief discussion of the history of the development of relativity see the entry "Einstein, Albert." This entry describes the content of the theories.

The special and general theories of relativity are, at heart, theories of spatiotemporal structure. They are not particularly about observers or reference frames or ways to synchronize clocks, although as fundamental physical theories they have implications about what observers will observe and what various physical procedures for coordinating clocks will accomplish. It is easy to fall under the impression that these theories are basically concerned about coordinate systems or reference frames because physical events are typically described by means of coordinates or reference frames, but that temptation ought to be avoided.

Perhaps the easiest way to understand special relativity is by analogy to Euclidean geometry. Euclidean geometry postulates a particular spatial structure and, beginning with the Euclid's *Elements*, the implications of that structure for geometrical figures were studied by purely geometrical methods. For two millennia, the study of Euclidean geometry made no use of coordinate systems or of numbers. The introduction of Cartesian coordinates allowed for the translation of geometrical objects into algebraic ones by means of assigning numbers as coordinates to points. There are all sorts of ways to lay down coordinates on a Euclidean space, such as polar coordinates or spherical coordinates, but the most familiar is the system of Cartesian coordinates. Cartesian coordinates are rectilinear and orthogonal; the coordinate curves are straight lines that intersect at right angles. Because of this feature, distances between points in a Euclidean space are easy to calculate from their Cartesian coordinates: If point  $p$  has coordinates  $(x_p, y_p, z_p)$  and point  $q$  has coordinates  $(x_q, y_q, z_q)$ , then the distance from  $p$  to  $q$  is:

$$\sqrt{(x_p - x_q)^2 + (y_p - y_q)^2 + (z_p - z_q)^2}.$$

In most spaces, such as the surface of a sphere, Cartesian coordinates do not exist. It turns out that for a space to be Euclidean is just for the space to admit of Cartesian coordinates. That is, the distances between points in the space must be of just the right form for Cartesian coordinates to exist.

In order to grasp relativity, we have to think not of distances between points in a three-dimensional space, but of a fundamentally spatiotemporal distance between

points in a four-dimensional space-time. Points in the space-time correspond to instantaneous, localized events, such as the bursting of a bubble when it reaches the surface of a glass of champagne. Such events occur both at a place and at a time. To locate these events, we typically ascribe to them four numbers, such as a latitude, longitude, altitude, and time. It is in this uncontroversial sense that the space-time of classical physics and of relativity is four-dimensional.

What sorts of spatiotemporal relations are there between events? All of classical physics agreed on at least one point: There is a definite, objective, purely temporal relation between the events. Two events either take place at the same time, or one takes place a certain amount of time before the other. So the notion of there being a lapse of time between events, and the specific case of simultaneity of events, is inherent in the classical account of space-time structure.

The classical account of spatial structure is not so straightforward. Newton believed that a single three-dimensional Euclidean space persists through time, and that every event, whenever it occurs, takes place somewhere in that absolute space. So Newton thought that any pair of events, no matter whether they occur at the same or different times, have some spatial distance separating them. But consider the following case: On a train traveling along the tracks, there sits a glass of champagne. A bubble rises to the surface and pops, followed a minute later by a second bubble. How far was the first popping from the second?

According to a passenger on the train, the two events took place in close spatial proximity, within a few inches of each other. But according to a spectator watching the train go by, these two events would be considered yards apart because the train has moved in the intervening minute. Newton would insist that there is a true spatial distance between the events, even though no observation could reveal for certain whether the passenger or the spectator (or neither) is right. But a natural reaction is to reject the whole question: There may be definite spatial relations between simultaneous events, but there is no fact at all about the spatial distance between nonsimultaneous events. Thus we arrive at two classical space-time structures: Newtonian space-time, with temporal and spatial relations between every pair of events, and Galilean (or neo-Newtonian) space-time, with temporal relations between all events and spatial relations only between simultaneous events (Galilean space-time then needs to add a new spatiotemporal structure, called an affine connection, to distinguish inertial from non-inertial trajectories). Note that the classical accounts agree on

the temporal structure, and particularly on the objective physical relation of simultaneity.

Special relativity postulates a four-dimensional space-time with a radically different spatiotemporal structure. Instead of having a pure temporal structure and a pure spatial structure, there is a single relativistic “distance” between events (the scare quotes around *distance* must be taken seriously, as the quantity is not at all like a spatial distance). How can this spatiotemporal structure be specified?

The easiest method, albeit a bit roundabout, is by means of coordinates. Here we will take the analogy with Euclidean geometry quite seriously. As we saw, even though Euclidean geometry has no need of coordinate systems, the spatial structure of a Euclidean space can still be specified in this way; a Euclidean space is a space that admits of Cartesian coordinates. More specifically, a three-dimensional Euclidean space has a structure of distance relations among its points such that each point can be given coordinates  $(x,y,z)$  and the distance between any pair of points is:

$$\sqrt{(x_p - x_q)^2 + (y_p - y_q)^2 + (z_p - z_q)^2}.$$

In exactly the same way, we can specify the spatiotemporal structure of Minkowski space-time, the space-time of special relativity. Minkowski space-time is a four-dimensional manifold that admits of Lorentz coordinates (or Lorentz frames). A Lorentz frame is a system of coordinates  $(t, x, y, z)$  such that the relativistic spatiotemporal distance between any pair of events  $p$  and  $q$  is:

$$\sqrt{(t_p - t_q)^2 - (x_p - x_q)^2 - (y_p - y_q)^2 - (z_p - z_q)^2}.$$

Written this way, the similarity with the example of Cartesian coordinates on Euclidean space is manifest; the only difference is the minus signs in place of plus signs. The consequences of that small mathematical difference are profound.

Before investigating the nature of this spatiotemporal structure, we should renew some of our caveats. First, there is always the temptation to invest the coordinates with some basic physical significance. For example, it is very natural to regard the coordinate we are calling  $t$  as a time coordinate, and to suppose that it has something to do with what is measured by clocks. But as of yet, we have said nothing to justify that interpretation. The Lorentz coordinates are just some way or other of attaching numbers to points such that the quantity defined above is proportional to the spatiotemporal distance between events. Indeed, just as there are many ways to lay down Cartesian coordinates on a Euclidean plane, systems differ with respect to the origin and orientation of the coordinate grid, so there are

many ways to lay down Lorentz coordinates in Minkowski space-time. Different systems will assign different  $t$  values to the points, and will disagree about, for example, the difference in  $t$  value between two events. We do not invest these differences with any physical significance; because the various systems agree about the quantity defined above, they agree about all that is physically real.

A second caveat is in order. We have been speaking so far as if the spatiotemporal distance between events is itself a number (viz., the number that results when one plugs the coordinates of the events into the formula above). But it is easy to see that this is wrong even in the Euclidean case. Distances are only associated with numbers once one has chosen a scale, such as inches or meters. What exists as a purely geometrical, nonnumerical structure is rather a system of ratios of distances. Having chosen a particular geometrical magnitude as a unit, other magnitudes can be expressed as numbers (viz., the numbers that represent the ratio between the unit and the given magnitude). The Greeks had a deep insight when they divided mathematics into arithmetic (the theory of number) and geometry (the theory of magnitude). They recognized that the theory of ratios applied equally to each field, but kept the two subjects strictly separate. Our use of coordinates to associate curves in space with algebraic functions of numbers has blurred the distinction between magnitudes and numbers. To understand relativity, it is important to recognize the conventions employed to associate geometrical structure with numerical structure.

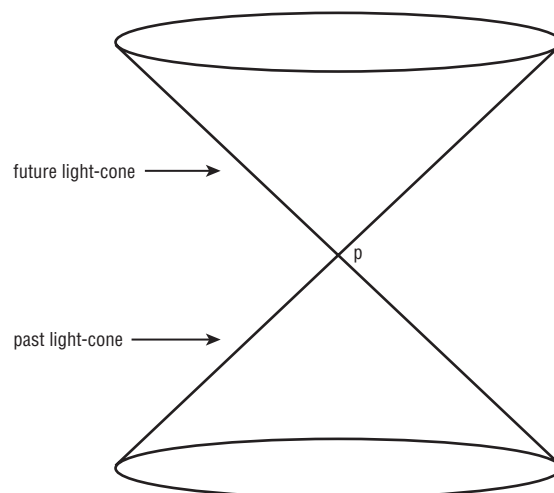
Holding these warnings in mind, let us turn to the relativistic spatiotemporal distance. What are the consequences of replacing the plus signs in the Euclidean distance function with minus signs?

One obvious difference between the Euclidean structure and the Minkowski structure is this: In Euclidean space, the distance between any two distinct points is always positive, and the only zero distance is between a point and itself. In mathematical terms, the Euclidean metrical structure is positive definite. But in the Minkowski structure, two distinct events can have zero distance between them. For example, the events with coordinates  $(0,0,0,0)$  and  $(1,1,0,0)$  have zero distance (where we list the coordinates in the order  $(t,x,y,z)$ ). Of course, this does not mean that these two events are the same event; assigning the numerical value zero to this sort of distance is just a product of the conventions we have used for assigning numbers to the distances. But the fact that two events have a zero distance between them does show that they are related in a particular spatiotemporal way. In order to remind ourselves that these spa-

tiotemporal distances do not behave like spatial distances, from now on we will call them spatiotemporal intervals.

If we choose a particular event, the popping of a particular champagne bubble, and call the event  $p$ , then we can consider the entire locus of events that have zero interval from  $p$ . There will be infinitely many such events. If  $p$  happens to be at the origin of a Lorentz frame, assigned coordinates  $(0,0,0,0)$ , then among the events at zero interval from it are  $(1,1,0,0)$ ,  $(1,0,1,0)$ ,  $(5,0,-3,4)$ , and  $(-6,4,-4,2)$ . To get a sense of how these events are distributed in space-time, we draw a space-time diagram, but again one must be very cautious when interpreting these diagrams. The diagrams must represent one or two dimensions of the space-time, because we cannot draw four-dimensional pictures, but that is not the principle problem. The main problem is that the diagrams are drawn on a Euclidean sheet of paper, even though they represent events in Minkowski space-time. There is always the danger of investing some of the Euclidean structure of the representation with physical significance it does not have. Bearing that in mind, the natural thing to do is to suppress the  $z$  coordinate and draw the  $x$ ,  $y$ , and  $t$  coordinates as the  $x$ ,  $y$ , and  $z$  coordinates of three-dimensional Euclidean space.

Adopting these conventions, the points at zero interval from  $(0,0,0)$  will be points that solve the equation  $t^2 - x^2 - y^2 = 0$ , or  $t^2 = x^2 + y^2$ . The points that solve this equation form a double cone whose apex is at the origin. According to relativity, the intrinsic spatiotemporal structure associates such a double cone with every event in the space-time. This locus of points is called the light-cone of the event  $p$ , and divides into two pieces, the two cones that meet at  $p$ . These cones are called the future light-cone and the past light cone of  $p$ .



As the name light-cone suggests, we are now in a position to make contact between the spatiotemporal structure postulated by relativity and the behavior of physical entities. According to the laws of relativistic physics, any light emitted at an event (in a vacuum) will propagate along the future light-cone of the event, and any light that arrives at an event (in a vacuum) arrives along the past light-cone. So the tiny flash of light emitted when our champagne bubble pops races away from the popping event along its future light-cone. One can think of the ever-growing light-cone as representing the expanding circle (or, if we add back the  $z$  dimension, the expanding sphere) of light that originates at the bursting of the bubble.

Having associated the spatiotemporal structure with the behavior of an observable phenomenon such as light, we can now see how relativistic physics gains empirical content. For example, it is an observable fact that any pair of light rays traveling in parallel directions in a vacuum travel at the same speed; one light ray in a vacuum never overtakes another. This is not, of course, how material particles behave. One spaceship traveling in a vacuum can overtake another, or one electron in a vacuum can overtake another, because where a spaceship or an electron goes depends on more than the space-time location of the origin and direction of its journey. Two electrons can start out at the same place and time and set off in the same direction but end up in different locations because they were shot out at different speeds. Their trajectories depend on more than just the space-time structure. Light, in contrast, is intimately and directly tied to the relativistic space-time structure. Space-time itself, as it were, tells light in a vacuum where to go.

The assignment of zero relativistic interval between the origin of a light-cone and any event on it has one other notable consequence. We have already said that when we assign numbers to magnitudes, we want the ratios between the numbers to be identical to the ratios between the magnitudes. Because  $0:0$  is not a proper ratio, the relativistic interval does not license comparisons between the various intervals on a light-cone. If one light ray originates at  $(0,0,0,0)$  and travels to  $(1,1,0,0)$ , and a second light ray originates at  $(0,0,0,0)$  traveling in some other direction, there is no fact about when the second light ray has gone as far as the first.

What other structure, beside the light-cone structure, does Minkowski space-time have? There is a well-defined notion of a straight line in the space-time, and this is accurately represented in our Euclidean space-time diagram: Straight lines in the Euclidean diagram corre-

spond to straight trajectories in the space-time. Indeed, we have tacitly been appealing to the notion of a straight line all along; when we speak of the relativistic interval between two events, we mean the interval as measured along a straight line connecting the events, or, even more precisely, we mean the relativistic length of the straight line that connects the events. The straight-line structure (affine structure) of Minkowski space-time plays a central role in framing physical laws.

If a light ray is emitted from  $(0,0,0,0)$  into a vacuum, we already know that its trajectory through space-time will lie on the future light-cone of  $(0,0,0,0)$ . But more than that, the trajectory will be a straight line on the light-cone. An analogous fact holds for material particles that travel below the speed of light. If a material particle is emitted from  $(0,0,0,0)$ , its trajectory will lie entirely within the future light-cone of  $(0,0,0,0)$ , which is to say that the particle can never travel at or above the speed of light. But more than that: If the particle is emitted into a vacuum, and is not subject to any forces, then its trajectory will be a straight line in space-time.

This law, in abstract form, enormously predates the theory of relativity. For this is just the proper space-time formulation of Newton's first law of motion: "Every body continues in its state of rest, or of uniform motion in a right line, unless compelled to change that state by forces impressed on it." The trajectory of a particle at rest or in uniform motion in Newtonian space-time is a straight line through the four-dimensional space-time. Newton's first law, stated in terms of space-time trajectories, also retains the same form in Galilean space-time, and can be taken over without change into Minkowski space-time. As we will see, in this abstract space-time formulation, Newton's first law also holds in the general theory of relativity. That is why we should try to formulate physical laws directly in terms of space-time structure.

Once we deal with material particles that travel below the speed of light, the relativistic interval takes on even greater significance. Consider a particle that travels from  $(0,0,0,0)$  to  $(5,4,0,0)$  along a straight trajectory (i.e., a particle emitted from the origin of the coordinate system that arrives at the event  $[5,4,0,0]$  without having any forces acting on it). The relativistic interval along its space-time trajectory is:

$$\sqrt{(5-0)^2 - (4-0)^2 - (0-0)^2 - (0-0)^2} = 3.$$

The size of this interval has direct physical significance; it is proportional to the amount of time that will elapse for a clock that travels along that trajectory. Clocks in the theory of relativity are like odometers on cars; they meas-

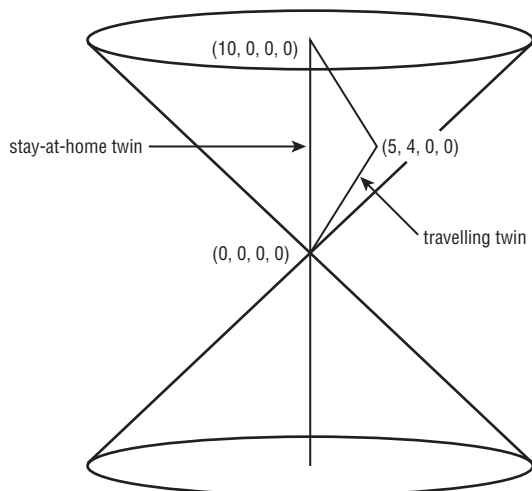


ure the length of the path they take. But *length* here means the interval, and *path* the space-time trajectory of the clock. Events in space-time separated by positive intervals are time-like separated.

It is not, of course, a further unanalyzable postulate of relativity that clocks measure the interval along their trajectory; clocks are physical mechanisms subject to physical analysis. But one can easily analyze how a simple clock will behave, such as a clock that counts the number of times a light ray gets reflected between two mirrors, and find that the reading on the clock will be proportional to the interval along the clock's trajectory.

With the clock postulate in hand, we can now analyze the notorious *twins paradox* of relativity. One of a pair of twins takes a rocket from Earth and travels to a nearby star. Upon returning to Earth, the twin has aged less than the stay-at-home sister, and the clocks in the twins' spaceship show less elapsed time than those that remained on Earth. Why is that?

To be concrete, suppose the event of the rocket leaving Earth is at the point  $(0,0,0,0)$  in our coordinate system, and the rocket travels inertially (without acceleration) to the point  $(5,4,0,0)$ . The rocket immediately turns around, and follows an inertial trajectory back to Earth, arriving at the event  $(10,0,0,0)$ . The interval between  $(0,0,0,0)$  and  $(5,4,0,0)$  is, as we have seen, 3. Suppose this corresponds to an elapse of three years according to the onboard clocks. The return trajectory from  $(5,4,0,0)$  to  $(10,0,0,0)$  also has an interval length 3, corresponding to another three years elapsed. So the astronaut twin arrives back having aged six years, and having had all the experiences that correspond to six years of life.



The stay-at-home twin, however, always remained at the spatial origin of the coordinate system. Her trajectory

through space-time is a straight line from  $(0,0,0,0)$  to  $(10,0,0,0)$ . So the interval along her trajectory is 10, corresponding to an elapse of ten years. She will have biologically aged ten years at her sister's return, and had four more years of experience than her twin.

The relativistic analysis of the situation is quite straightforward. It is really no more surprising, from a relativistic perspective, that the clocks of the twins will show different elapsed times from departure to return than it is surprising that two cars starting in the same city and end in the same city will show different elapsed mileage on their odometers, given that one took the free-way and the other a winding scenic route. The sense that there is a fundamental puzzle in the twins paradox only arises if one has mistaken views concerning the content of the theory of relativity.

In particular, it is often said that, according to the theory of relativity, all motion is the relative motion of bodies. If so, then there seems to be a complete symmetry between the twins: The motion of twin A relative to twin B is identical to the motion of twin B relative to twin A. But the relative motion of the twins plays no role at all in the physical analysis of the situation. The amount of time that elapses for twin B on her trip has nothing to do with what twin A is doing, or even if there is a twin A. The amount of time is just a function of the space-time interval along her trajectory.

It is also sometimes said that the theory of relativity gets rid of all absolute spatiotemporal structure; all facts about space and time are ultimately understood in terms of relations between bodies, so in a world with only one body there could be no spatiotemporal facts. This is also incorrect. The special theory of relativity postulates the existence of Minkowski space-time, whose intrinsic spatiotemporal structure is perfectly absolute, in whatever sense one takes that term. It is not a classical space-time structure, but it is not just a system of relations between bodies.

One occasionally also hears that the resolution of the twins paradox rests on facts about acceleration; the situation of the two twins is not exactly symmetric because the astronaut twin must accelerate (when she turns around to come home), whereas the stay-at-home twin does not. That is true, but irrelevant: The difference in elapsed time is a function of the intervals along the trajectories, not a function of the accelerations that the twins experience. Indeed, in the general theory of relativity we will be able to construct a twins scenario in which neither twin accelerates at all, but still they suffer different elapsed times between parting and reunion. It would be just as mis-

leading to attribute the difference in elapsed time to the accelerations of the twins as it would the difference in odometer reading to the accelerations of the cars, even if the car that took the longer route did accelerate more.

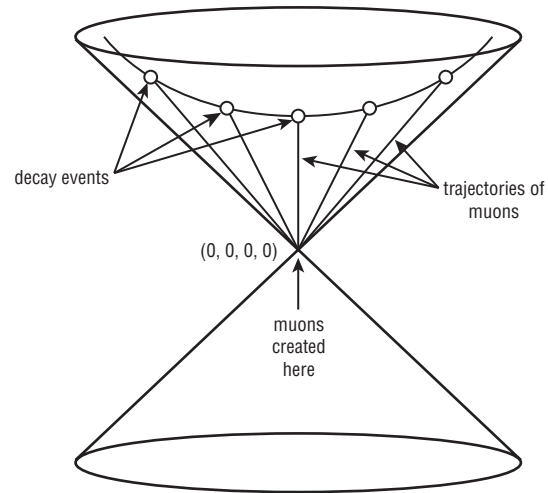
The paradoxical or puzzling aspect of the twins paradox really arises from the difference between Euclidean geometry and Minkowski space-time geometry. If we draw the trajectories of the twins in space-time, we get a triangle whose corners lie at  $(0,0,0,0)$ ,  $(5,4,0,0)$ , and  $(10,0,0,0)$ . The astronaut twin travels along two edges of this triangle, whereas the stay-at-home twin travels along the third. And in Euclidean geometry, the sum of the lengths of any two sides of a triangle are greater than the length of the remaining side. But in Minkowskian geometry, the opposite is true: The sum of the intervals of two sides is less than the interval along the remaining side. Indeed, for time-like separated events, a straight line is the longest path between the two points in space-time. This is one consequence of exchanging the plus signs in the Euclidean metric for minus signs in the Minkowski metric.

The relativistic clock postulate has been most strikingly checked using natural clocks: unstable particles whose decay rate displays a known half-life in the laboratory. The muon, a sort of heavy electron, is unstable and will decay on an average of  $10^{-6}$  seconds after having been created. Muons can be created in the upper atmosphere by collisions between molecules in the air and high-energy cosmic rays. According to clocks on Earth, it should take the muon about  $10 \times 10^{-6}$  seconds to reach the Earth, so very few should survive the trip without decaying. Nonetheless, many more muons than that calculation suggests do reach the Earth's surface. Calculation of the interval along the muon's trajectory predicts this because that interval corresponds to less than  $10^{-6}$  seconds.

If we idealize muons a bit, and imagine that they all decay in exactly  $10^{-6}$  seconds (according to their own clocks), then we can use them to map out the geometry of Minkowski space-time. Suppose we create a swarm of muons in space and send them out in all directions. Their decays will provide a map in space-time of events that are all the same interval from the point of creation. If we choose units so that the size of the interval corresponds to seconds, and we choose the creation of the muons as the origin of the coordinate system, then the coordinates of the decay events will satisfy:

$$\sqrt{t^2 - x^2 - y^2 - z^2} = 10^{-6}$$

This is the equation of a hyperboloid of revolution that asymptotically approaches the light-cone, as depicted below.



The hyperboloid represents events all at the same interval from  $(0,0,0,0)$ , and so corresponds to a circle or sphere of fixed radius in Euclidean geometry. There would be a corresponding hyperboloid in the past light-cone, representing places from which a muon could have been sent that would have decayed at  $(0,0,0,0)$ .

Indeed, we are now in a position to make a thoroughgoing analogy between the geometry of Minkowski space-time and Euclidean geometry that makes no reference to coordinates at all. Classical Euclidean geometrical proofs do not use coordinate systems of numbers, they use two instruments: the straightedge and the compass. The straightedge allows one to identify straight lines in the space, and the compass to draw the locus of points at a fixed distance from a given center. In Minkowski space-time, we can use light rays in a vacuum and inertially traveling particles as straightedges because their trajectories are straight lines in the space-time. Setting a Minkowski compass at interval zero and identifying a center should result in drawing the light-cone: the locus of points of interval zero from the center. So we can use light rays for this purpose. Setting the compass to draw points at a fixed positive interval should result in drawing hyperbola; we can use clocks for this just as the muons are employed above. In this way, we can free Minkowski geometry from coordinates altogether.

So far we have left one species of space-time relation out of account. All the points on the past or future light-cone of some event  $p$  are at zero interval from  $p$ . All the events inside the past or future light-cone are at positive interval from  $p$  (taking always the positive square root by

convention). What of points that are outside the light-cone altogether?

The point labeled (0,1,0,0) is outside the light-cone of the point (0,0,0,0). If we plug these coordinates into our formula, we find that the interval between the points is:

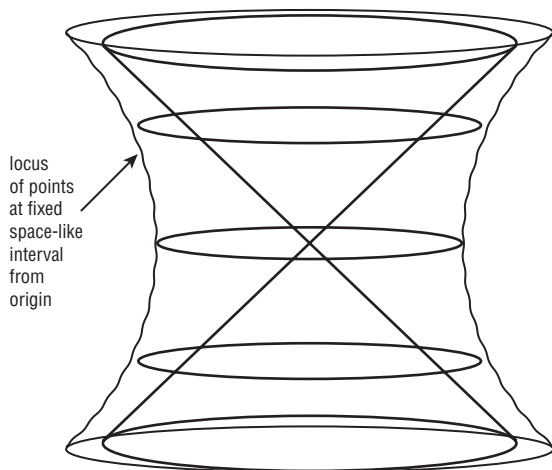
$$\sqrt{(0-0)^2 - (1-0)^2 - (0-0)^2 - (0-0)^2} = \sqrt{-1} = i.$$

That is, according to the definition of the interval that we have given, the interval between these points is imaginary. What could this mean?

Once again, we have to recall that the assignment of numbers to the intervals is somewhat a matter of convention. In fact, some physics books define the interval as:

$$\sqrt{(x_p - x_q)^2 + (y_p - y_q)^2 + (z_p - z_q)^2 + (t_p - t_q)^2},$$

Here the interval between time-like separated events becomes imaginary. Does this mean that a clock could measure an imaginary number? Of course it can: Just take a regular clock and paint a little  $i$  after all the numerals! The numbers we assign to intervals have no intrinsic significance; it is the ratios between the numbers that represent the ratios among the magnitudes. Events that lie outside each other's light cones, so-called space-like separated events, have intervals among them that also stand in ratios to each other. The set of events at fixed space-like separation from (0,0,0,0) forms another sort of hyperboloid of revolution, depicted below.



We now have a sense of the spatiotemporal structure of Minkowski space-time. A special relativistic physical theory must have laws that employ only this spatiotemporal structure. We could now go on to see how, for example, classical electromagnetic theory can be reformulated in this way, but that would take us too far from foundational issues.

It should be noticed that this account of special relativity has made no mention at all of several well-known features often associated with relativity, such as the constancy of the speed of light, the relativity of simultaneity, and the Lorentz-Fitzgerald contraction. That is because all of these are frame-dependent (or coordinate system dependent) effects, and we have been presenting the theory in a frame-independent way. For example, we have no basis to discuss the relativity of simultaneity because we have had no ground, and no need, to introduce any notion of simultaneity at all. In classical physics, simultaneous events are events that take place at the same time, but we have no general notion of the time at which an event occurs, only the time that elapses on a clock following a certain trajectory. So the proper thing to say is not that special relativity implies the relativity of simultaneity, but that it implies the nonexistence of any objective notion of simultaneity. And we cannot discuss whether the speed of light is constant because we do not have any grounds to ascribe any speed to anything.

We have seen that a light ray can never overtake another light ray, but assessing a speed requires determining how far an object went in a given period of time. So far, we have not needed any notion of the distance an object travels, nor of the time that it takes to travel that distance. We can say how much time will elapse on a clock that follows a given trajectory, but that is evidently no use in defining a speed of light; no material clock can travel along with a light ray, and if it could, it would show no elapsed time for the journey. The notion of simultaneity requires a global time function, that is, an assignment of times to all events, so that there is a locus of events that are all assigned the same time. And the notion of a speed requires both the notion of the time that elapses between the start and the end of a journey, and the notion of the distance covered in that time. The relativistic space-time structure does not, per se, support either of these notions.

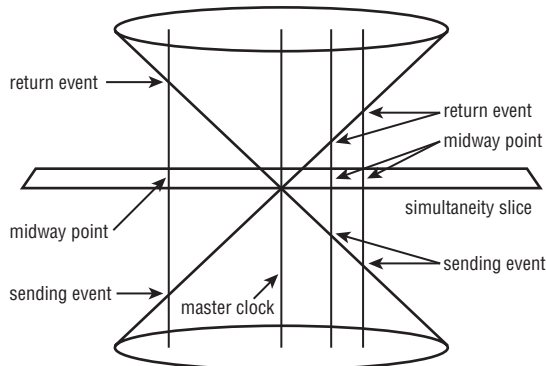
There is, however, a reasonably natural method for introducing both a global time function and a notion of spatial distance into Minkowski space-time. We begin with a family of co-moving inertial clocks (i.e., a family of clocks all moving on straight, parallel trajectories through space-time). There will be an infinitude of such families, corresponding to all the directions their trajectories can have. We begin by picking one such family.

We now want to “synchronize” the clocks. Scare quotes have to be put around the word since the classical notion of synchronization presupposes the notion of simultaneity: Synchronized clocks all show the same time at the same instant. But in relativity there is no such thing

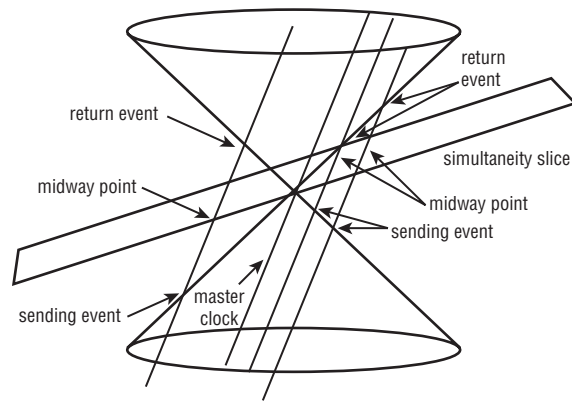
as *the same instant*. So one must think of the method we are about to describe as a way to coordinate a family of clocks that we simply call synchronizing them.

Let us choose a single master clock from our family of co-moving clocks. The other clocks will coordinate with this master clock by the following method: Each clock sends a light ray to the master clock, noting the time of emission (according to the sending clock). When the light ray reaches the master clock, it is immediately sent back and shows the time reading on the master clock at the moment it arrived. When this return signal reaches the sending clock, the time reading on the sending clock is noted. The sender, then, has three bits of data: the time it sent the signal (according to the sending clock), the time it received the return signal (ditto), and the reading on the master clock when the signal got to it. On this basis, the sending clock synchronizes with the master clock by adjusting its time so that the time that the master clock read when the signal arrived corresponds to the event on the sending clock exactly midway between the moment the signal was sent and the moment the return signal arrived. All of these notions are relativistically well-defined, so this method of coordinating clocks can be carried out. Every event in space-time is now assigned a time (viz., the reading on that member of the family of clocks that passes through the event when it passes through the event).

We can now identify simultaneity according to this family of clocks as sets of events that are all assigned the same time by this family of clocks. Such a set is called a simultaneity slice through the space-time. The figure below shows one such simultaneity slice. Because all of the light signals that reach the master clock at noon lie on the past light-cone of the master clock showing noon, and because all of the return signals lie on the future light-cone of that event, it is easy to calculate the points at which all of the coordinated clocks will register noon. It is the flat plane in the middle.



The simultaneity slice is a function of which family of co-moving clocks we choose. Choosing another family will give a different notion of simultaneity:



Each family of co-moving clocks determines its own notion of simultaneity, and these various notions render different judgments concerning which pairs of events happen at the same time. All the families will agree about the time order of time-like or light-like separated events, but for any pair of space-like separated events, some families will say that they happened at the same time, others that one happened first, and yet others that the other happened first. Each family introduces its own global time function. None of these functions is superior to the other, and none is needed at all to explicate the basic spatio-temporal structure.

What of spatial distance? Once a family of clocks has been synchronized, there is a simple way to assign a spatial distance between any pair of clocks. Send a light ray from one clock to the other. We can now understand the time of travel for the light ray as the difference between the time showing on the emitting clock at the emission event and the time showing on the receiving clock at the reception event. So we now have a definition of how long the light ray took to get from one clock to the other (again, this is not the time that a clock traveling along with the light ray would show elapsing). If we now define the speed of light to be a given constant,  $c$ , then we can say that the distance between the clocks is just  $c$  times the elapsed time of transmission. This will give us a structure of spatial distances between the clocks as defined by that particular family of clocks. Those spatial distances will, in special relativity, constitute a Euclidean space. Different families of clocks will disagree about the precise spatial distance between events, and about the spatial size of material objects, but each family will construct for itself a Euclidean spatial structure. Finally, if we allow such a

family of clocks to introduce Cartesian coordinates on its Euclidean space, then the family will assign each event four coordinate numbers: the three spatial coordinates and the global time function. These are exactly the Lorentz coordinate frames that we began with to express the relativistic metric, so we have come full circle.

The interconnection between the global time defined by a family of clocks and the spatial structure among events defined by that family resolves many of the intuitive puzzles in special relativity. We have seen that, according to clocks at rest on the Earth, a high-energy muon has a much longer lifetime than a muon at rest. That explains, from the point of view of the Earth frame, how the muon manages to make the trip to the surface. But of course, from the point of view of the muon, and clocks co-moving with it, the muon lifetime is the normal 10<sup>-6</sup> seconds. From their point of view, the Earth is approaching them at high velocity. In that frame of reference, the muon is able to get through the whole atmosphere not because of any slowing down of their clocks, but because of the spatial contraction of the atmosphere. In the muon's frame of reference, the distance from the upper atmosphere to the Earth is much less than we on Earth take it to be.

The Lorentz contraction and time dilation effects of relativity then arise as disagreements that occur between the Lorentz frames about the amount of time that elapses between events and the spatial distance between events. Clocks in any frame will be seen to run slow according to the time function associated with any other frame. A meter stick at rest in one frame will be judged to be less than a meter long according to a frame in which the stick is moving. These are symmetric effects: From the point of view of any Lorentz frame, clocks at rest in any other frame run slow. We need to sharply distinguish these effects from the twins paradox. There, the difference in elapsed time for each twin is a consequence of the fundamental spatiotemporal structure, and has nothing to do with frames or families of clocks. The time dilation between frames results only from different ways of defining coordinates. In the latter case, there is no fact about which set of clocks is really running slower, but in the former case there is an objective fact about which twin is biologically younger when they are reunited.

## GENERAL RELATIVITY

Special relativity is a theory that postulates a certain intrinsic spatiotemporal structure, and then formulates the laws of physics in terms of that structure. General relativity is the relativistic theory of gravity. It is also funda-

mentally a theory about spatiotemporal structure, and allows for different structures than special relativity. So the first question that arises when approaching general relativity is why gravity should particularly be connected to spatiotemporal structure. The special relativistic theory of electromagnetism, for example, simply accepts the Minkowski space-time and employs it in framing the electromagnetic laws. But gravity, in contrast, led to the rejection of special relativity in favor of a new theory. What is so special about gravity?

One sometimes hears that there needed to be a relativistic theory of gravity because Newton's gravitational theory postulates that gravity acts instantaneously between distant masses, but in relativity there is no available notion of instantaneous action (because there is no physical notion of simultaneity). But this observation does nothing to suggest that the theory of gravity should require any change from the special relativistic space-time. Classical electrostatics postulated that the coulomb force between distant charged particles acts instantaneously, but electromagnetic phenomena do not require changes to special relativity. Rather, relativistic electrodynamics simply rejects the claim that electric and magnetic forces act instantaneously. Electromagnetic influences are propagated along the light cones, at the speed of light, by electromagnetic waves. Similarly, one might think that the obvious way to deal with gravitation is simply to deny that it acts instantaneously. Let the gravitational effects also propagate along the light cones, and the special relativistic structure can be used to formulate the laws.

Adding such a delay in gravitational influence would, of course, modify the predictions of Newtonian gravity. One might even plausibly argue that Newton himself would have expected such a correction to his instantaneous gravity. For Newton thought that gravitational forces were mediated by particles exchanged between the gravitating bodies, and he would have expected the particles to take some time in traveling between the bodies. Of course, the fundamental cause of the gravitational force was a topic on which Newton refused to *fingere* any hypothesis, so we must be a bit speculative here. But it is worthwhile to note that if we modify classical Newtonian gravitational theory to allow gravitational influence to propagate along the light cones, we can exactly derive some famous relativistic effects, such as the anomalous advance in the perihelion of Mercury.

In order to understand why gravity is plausibly taken to be deeply connected to space-time structure, we need to look elsewhere. Consider again the family of co-moving inertial clocks we made use of in our discussion

of special relativity. Once set in motion, the family of clocks will move together, never approaching or receding from each other. That is because: a) the clocks are all traveling inertially, not subject to any force; b) according to the space-time version of Newton's first law, the trajectories of bodies subject to no forces will be straight lines in space-time; and c) the straight-line trajectories of the co-moving clocks form a family of parallel straight lines. Note that in giving this argument, we never had to mention the mass of any of the clocks. Because they are moving inertially, the trajectories of the clocks are determined by the intrinsic space-time structure, without the mass playing any role. It would not matter if some of the clocks were heavy and others light; they would still move together parallel to one another.

In Newtonian physics, the mass of a body only comes into consideration when a body is subject to a force and thereby deflected off its inertial trajectory. The inertial mass of a body is nothing but a measure of the body's resistance to being deflected by a force from its inertial trajectory: The more massive a body is, the harder it is to make its trajectory bend in space-time. Newton's second law, which we now render  $\mathbf{F} = m\mathbf{A}$ , tells us that the same force will only produce half the acceleration in a body that is twice as massive. So in the presence of forces, the trajectories of bodies will depend on their masses, whereas in the absence of forces the more and less massive bodies will move on parallel trajectories. Turning this observation around, we should find it very suggestive if there is a situation in which the trajectory of a body does not depend at all on its mass. It is natural to suspect that in such a situation, the mass of the body is playing no role because the body is not being subject to any force; it is moving inertially.

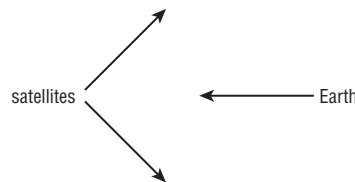
Recall Galileo at the top of the Leaning Tower of Pisa dropping a lighter and heavier object and seeing them hit the ground together. Here is a common situation in which the mass of an object does not affect its trajectory: The heavy and light follow the same space-time path. According to Newtonian gravitational theory, this is a rather fortuitous result. In that theory, both the heavy and the light object are subject to a force, the force of gravity, and so each is being deflected off its inertial trajectory. But, luckily, the gravitational force on each object is exactly proportional to its inertial mass. So the more massive object, which needs a greater force to be accelerated, is subject to a greater force than the less massive object. Indeed, the gravitational force on the more massive object is exactly as much larger as it needs to be to produce precisely the same acceleration as the lesser force

of gravity produces on the less massive object. That, according to Newton, is why they fall together; they are both accelerated, but at exactly the same rate.

If we follow the hint above, though, we will be led to suspect a different account. Perhaps the two objects move together not because they are equally deflected off their inertial, straight-line trajectories, but rather because they are both following their inertial trajectories. Because the inertial trajectories are straight lines in space-time, this suggests a deep connection between gravity and fundamental spatiotemporal structure.

In this way we arrive at the general theory of relativity. According to general relativity, objects that are falling in a gravitational field or under the influence of a gravitational force are not being affected by any force at all. Gravity does not deflect objects from their inertial paths, it rather influences the very structure of space-time itself. The balls falling from the Leaning Tower of Pisa, or the planets orbiting the sun, are following straight trajectories through space-time.

To realize this theory, we must reject Minkowski space-time. Consider, for example, two satellites orbiting the Earth in opposite senses. The space-time diagram of the situation looks like this:



As the satellites orbit, their paths cross and recross in space-time. But in Minkowski space-time, as in Euclidean space, two straight lines can intersect only once at most. So the space-times of general relativity must have a different spatiotemporal structure than the space-time of special relativity.

An analogy with pure spatial geometry helps here. Euclidean geometry is just one of an infinitude of spatial geometries. Lines on the surface of a sphere, for example, do not satisfy Euclid's postulates. But even spherical geometry is highly regular compared to most geometries. Consider, for example, the surface of North America. In regions of the Great Plains, the geometry is nearly Euclidean (and even more nearly spherical), whereas in the Rocky Mountains the geometry of the surface varies wildly from place to place. We need new mathematical machinery to deal with this sort of situation.

The general mathematics needed is called differential geometry. Differential geometry is suited to deal with spaces whose geometrical structure varies from place to place. In some regions, a space may be locally Euclidean, in others non-Euclidean, so we have to be able describe the geometry region by region.

Euclidean spaces have a particularly uniform geometrical structure that allows them to admit of very convenient coordinate systems. As we have seen, a Euclidean space admits of Cartesian coordinates, in which the distances between points is a simple mathematical function of the coordinates of the points. Non-Euclidean spaces do not admit of such convenient systems. For example, points on a sphere can be coordinatized by latitude and longitude, but distances between the points on a sphere are not a simple function of their coordinate differences. If you are near the North Pole, you can change your longitude by several degrees just by taking a few steps; near the equator the same change of longitude would require traveling hundreds of miles. And even spherical coordinates are relatively simple and uniform.

To get a sense of a completely generic coordinate system, imagine walking down a road where each successive house has an address—one greater than the house before. You want to get to house number 200 and you are currently at house 100. How far must you walk?

There is no way to tell. If you go through a densely populated area, such as a small town, you will get to your destination quickly. If it is a sparsely built region, you may have to walk a long way. To know how far you have to go, you would need a complete listing of the distances between successive houses. If you have such a list, you can calculate the distance between any two houses, and so can reconstruct the geometrical structure of the region where the houses are built. In an analogous way, the general theory of spaces allows for the use of any arbitrary coordinate system. Accompanying the system is a metric that specifies the distances between nearby points. We do not have any general rule for calculating distances between distant points as a function of their coordinates, but we do not need one. The distance between faraway points is just the length of the straight path that connects them, and we can calculate the length of that path by knowing the distance between nearby points and adding up all the distances along the path. Thus we have the mathematical tools to deal with generic spaces of variable curvature that admit of nothing like Cartesian coordinates.

It is sometimes said that the general theory of relativity requires us to replace Euclidean space with a non-Euclidean space, but that is not a very useful, or accurate,

explanation of the situation. As we have seen, even in special relativity the notion of spatial geometry is rather derivative and non-fundamental. The fundamental notion is the relativistic interval, which is a spatiotemporal object. It is only relative to a family of co-moving objects, such as clocks, that we can even define a spatial geometry. It turns out that, in special relativity, each such family will ascribe Euclidean geometry to its space, but that is somewhat fortuitous; there is no logical guarantee that the various families will agree on their findings. After all, in special relativity the various families will disagree about the exact spatial distance (and temporal gap) between a given pair of events. In general relativity, there will, in general, not exist families of co-moving inertial observers that maintain the same spatiotemporal relations to one another, and so there is no unproblematic way to define a spatial geometry at all.

In any case, it is simply incorrect to say that objects moving in a gravitational field trace out straight paths in a non-Euclidean spatial geometry. The orbits of the planets, for example, are nearly elliptical in any reasonably defined space for the solar system, and the ellipses are not (spatially) straight lines.

The proper account of general relativity rather employs an analogy. As the variably curved non-Euclidean spaces are to Euclidean space, so the variably curved space-times of general relativity are to Minkowski space-time. The orbits of the satellites depicted above are not straight paths in any spatial geometry, but they are straight paths in space-time. The effect of the Earth is not to produce a force that deflects the satellites off their inertial paths, it is to alter the space-time geometry so that it contains inertial paths that cross and recross.

On the Newtonian picture of gravity, when we sit on a chair we are not accelerated because we are acted on by counterbalancing forces: The gravitational force pulling us down and the force of the chair pushing us up. According to the general relativistic account, the force of the chair pushing us up still exists, but it is unbalanced by any gravitational force. It follows that according to general relativity, as we sit we are constantly accelerated (i.e., constantly being deflected off of our inertial, straight-line trajectories through space-time). The inertial trajectory is that of an object unsupported by anything like the chair (i.e., an object in free fall).

The curvature of general relativistic space-time is partially a function of the distribution of matter and energy; that is why space-time near a massive object like the Earth is curved in such a way as to produce a gravitational field. This connection between the matter and

energy distribution and the spatiotemporal geometry is provided by Einstein's general relativistic field equations. But although the distribution of matter and energy influences the space-time geometry, it does not completely determine it.

The situation is similar to the relationship between the electromagnetic field and the electric charge distribution in classical physics. The presence of electric charges contributes to the electromagnetic field, but does not, by itself, determine it. For example, even in a space devoid of electric charges, there can be a nonzero electromagnetic field: electromagnetic waves (i.e., light) can propagate through the vacuum. Similarly, the general theory predicts the existence of gravitational waves—disturbances of the spatiotemporal geometry that can exist even in the absence of any matter or energy and that propagate at the speed of light. There are, for example, many vacuum solutions of the Einstein field equations. One solution is Minkowski space-time, but other solutions contain gravitational waves.

Because general relativity concerns spatiotemporal structure, and because the trajectory of light rays is determined by the light-cone structure, general relativity must predict the gravitational bending of light. It is not clear whether Newtonian physics would predict any gravitational effect on light because that would depend on whether light feels any gravitational force, but light certainly does propagate through space-time. The effect of gravity on light was dramatically confirmed in Arthur S. Eddington's 1919 eclipse expedition, but is even more strikingly illustrated in the phenomenon of gravitational lensing: A galaxy positioned between the Earth and a more distant light source can act as a lens, focusing the light of the distant source on the Earth. Two astronauts traveling inertially could experience a similar effect; they could take different straight paths that both originate at their home planet and both end on Earth, going different ways around an intervening galaxy. Because the relativistic interval along those paths could differ, such astronauts could illustrate the twins paradox without any acceleration; twins coming from the distant planet could have different biological ages when they reunite on Earth, even though neither suffered any acceleration.

The spatiotemporal geometry of general relativity accounts for familiar gravitational phenomena, but the theory also has dramatic consequence at the cosmological scale and in extreme physical conditions. When a massive star burns through its nuclear fuel and collapses, for example, the increasing density of matter causes ever greater curvature in space-time. If the star is sufficiently

massive, the light-cone structure deviates enough from Minkowski space-time to form a trapped surface: a region from which light cannot escape. The event horizon around a black hole is such a trapped surface; an object falling through the horizon can never send light, or any other signal, back to the exterior region. Once the infalling matter of the star reaches this point, it is destined to become ever more compressed without limit, and the curvature of the space-time will grow to infinity. If the equations continue to hold, this results in a space-time singularity; the spatiotemporal structure cannot be continued beyond a certain limit and space-time itself comes to an end. Because the spatiotemporal structure itself has become singular, it no longer makes any conceptual sense to ask what happens after the singularity; no meaning could be attached to the term after in the absence of spatiotemporal structure.

In the opposite temporal direction, the general theory also contains models in which the universe as a whole arises out of such a singularity, the singularity we call the big bang. Indeed, if general relativity is not modified, the observed motions of galaxies require that the universe began at a singularity, and that space-time itself has been expanding ever since. There is equally no sense to be made of the question what happened before the big bang because the spatiotemporal structure needed to define temporal priority would not extend beyond the initial singularity.

It is, of course, possible that the equations of the theory will be modified in some way so as to avoid the infinities and singularities, but that takes us from the analysis of general relativity into speculations about the replacement of general relativity. The mathematical structure of general relativity also admits of models of the theory with very peculiar spatiotemporal structures. Some models, for example, admit closed time-like curves, that is, time-like trajectories that loop back through space-time and meet up with themselves. In such a model, a person could in principle continue going always locally forward in time, but end up (as an adult) back at the events of their childhood. There seems to be no way to physically test this possibility (that is, there is no physical mechanism to produce closed time-like curves through laboratory operations), so it is unclear whether the existence of these mathematical models proves the physical possibility of such time travel or rather the physical impossibility of space-times that correspond to these mathematical solutions. In any case, general relativity provides a means for considering spatiotemporal structures unlike any that occur in classical physics.



The special and general theories of relativity provide a rich source of novel concepts of great interest to metaphysics. The topics that could be informed by these theories are too long even to list, but the most obvious metaphysical implications of the theories are worthy of remark. The nature of space and time occupies a central place in Immanuel Kant's *Critique of Pure Reason*, where supposed a priori knowledge of spatial and temporal structure provided grounds for the conclusion that space and time have no existence outside the faculty of intuition. After all, how could one know anything a priori about space and time if they exist outside the mind?

The theories of relativity simply refute the claim that there is any a priori knowledge of spatiotemporal structure. Even if relativity ultimately proves to be incorrect, everything in our everyday experience of the world can be accommodated in the relativistic spatiotemporal account. For all we know at present, we could be living in a relativistic universe, in which there is no Euclidean space and in which even time need not have a universal linear order. The nature of space and time is a matter of empirical inquiry, not a priori proof.

The special and general theories are also relevant to the question of the nature of space and time: Are they entities in their own right (as Newton supposed) or just relations among material bodies (as G.W. Leibniz insisted)? Taken at face value, the theories posit an independent existence to the four-dimensional space-time manifold. Even in the absence of material bodies, there is a spatiotemporal structure among the points of space-time. As the twins paradox shows, the observable behavior of material objects is determined by that structure. And even more dramatically, in general relativity the space-time manifold takes on a life of its own; gravitational waves can exist even in the absence of any material objects, and the presence of material objects influences the structure of space-time.

Attempts have been made to reformulate general relativity in a more relationist manner, in terms only of relations among material objects without commitment to any spatiotemporal structure of vacuum regions. These attempts have not succeeded. One can, of course, simply declare that in the general theory, space-time itself counts as a material entity, but then the argument seems to be only over labels rather than ontology.

Like all empirical theories, relativity is supported but not proven by observation. The spatiotemporal structure cannot be directly observed, but theories of matter couched in terms of the relativistic structure yield testable predictions that can be checked. The general theory, for

example, has been checked by flying an atomic clock around the world and comparing its reading with an initially synchronized clock that remained on Earth. Because the trajectories of the clocks have different relativistic intervals, one can predict that the traveling clock will show a different elapsed time from the clock that remained behind—which it does. There may be other ways to explain the effect, but it is a natural consequence of the relativistic account of space-time structure.

Challenges to the theory of relativity are more likely to come from considerations of the compatibility of the theory with other fundamental physical theories than from direct empirical problems. It is, for example, a still unsolved problem how to reconcile quantum physics with the pure relativistic space-time structure, and another unsolved problem of how to produce a quantum theory of gravity. Most particularly, the observable violations of John Bell's inequality for events at space-like separation are difficult to account for in any theory that has no preferred simultaneity slices in its space-time. So the metaphysician ought not to take the account of space-time provided by relativity as definitive; progress in physics may well demand radical revision of the account of spatiotemporal structure. Still, relativity illustrates how empirical inquiry can lead to the revision of the most seemingly fundamental concepts, even those that were once taken as preconditions for any experience at all.

**See also** Bell, John, and Bell's Theorem; Eddington, Arthur Stanley; Einstein, Albert; Energy; Galileo Galilei; Geometry; Knowledge, A Priori; Laws, Scientific; Leibniz, Gottfried Wilhelm; Matter; Motion, A Historical Survey; Newton, Isaac; Philosophy of Physics; Quantum Mechanics; Space; Time.

### **Bibliography**

- Bell, John S. "How to Teach Special Relativity." In *The Speakable and Unsayable in Quantum Mechanics*. Cambridge, U.K.: Cambridge University Press, 1987.
- Einstein, Albert. "On The Electrodynamics of Moving Bodies." In *The Principle of Relativity*. New York: Dover, 1952.
- Einstein, Albert. "The Foundations of General Relativity." *Annalen der Physik* 49 (1916).
- Geroch, Robert P. *General Relativity from A to B*. Chicago: University of Chicago Press, 1978.
- Maudlin, Tim. *Quantum Non-Locality and Relativity: Metaphysical Intimations of Modern Physics*. Oxford: Basil Blackwell, 1994.
- Misner, Charles, Kip S. Thorne, and John Archibald Wheeler. *Gravitation*. San Francisco: W. H. Freeman, 1973.
- Taylor, E. F., and J. A. Wheeler. *Spacetime Physics: Introduction to Special Relativity*. New York: W. H. Freeman, 1992.

**Tim Maudlin (2005)**

## RELEVANCE (RELEVANT) LOGICS

The conditional, “if ... then ...” ( $\rightarrow$ ) has been a contentious topic throughout the history of Western logic, and numerous accounts of its behavior have been proposed. One recurrent account (usually called the *material conditional*) is that  $A \rightarrow B$  is true just if the antecedent,  $A$ , is false or the consequent,  $B$ , is true. This account was built into the logic of Frege and Russell, and so came to assume orthodoxy throughout much of the twentieth century (at least where there are no subjunctive moods in the antecedent or consequent). The account has obvious problems, however. It entails, for example, that both of the following are true—which they do not appear to be: “If Melbourne is the capital of Australia, it is in China” (false antecedent), “If there is life on another planet, Canberra is the capital of Australia” (true consequent).

It is natural to suppose that in a true conditional the antecedent must be relevant to the consequent in some way. This idea is packed into the contemporary definition of a relevant logic. A propositional logic is a relevant/relevance (both words are used) logic just if whenever  $A \rightarrow B$  is a logical truth  $A$  and  $B$  share a propositional parameter (variable). (A quantifier logic is relevant if its propositional part is.)

Relevant logics can be of several different kinds. However, one has come to dominate current work in the area. This is the Anderson/Belnap tradition. Axiomatizations of logics (or fragments of logics) of this kind were proposed by Ivan Orlov (1928), Alonzo Church (1951), and Wilhelm Ackermann (1956). But the subject took off with the work of the Pittsburgh school of Alan Anderson and Nuel Belnap in the 1960s and 1970s. Probably the most important system of relevant logic developed by the school was the logic  $R$  (though Anderson and Belnap themselves preferred the system  $E$ ). This contained most of the intuitively correct principles concerning the conditional, but not “paradoxes” such as  $(A \& \neg A) \rightarrow B$  and  $A \rightarrow (B \rightarrow B)$ .

Semantics of various kinds for relevant logics were produced about ten years later by, among others, J. Michael Dunn, Alasdair Urquhart, and Kit Fine. But perhaps the most versatile semantics for relevant logics are the world-semantics developed by the Canberra school of Richard Sylvan (né Routley) and Robert Meyer (who had also been a member of the Pittsburgh school).

The world-semantics of relevant logics may be thought of as extending the possible-world semantics of

modal logic by adding a class of logically impossible worlds—though validity is defined in terms of truth-preservation at just the possible worlds. (This comes out most clearly in the simplified form of the semantics, as later developed by Graham Priest, Sylvan, and Greg Restall.) At a possible world,  $w$ , the truth conditions for  $\rightarrow$  are the same as those for the strict conditional in the modal logic  $S5$ :

$A \rightarrow B$  is true at  $w$  iff for all worlds,  $x$  (possible and impossible), when  $A$  is true at  $x$ ,  $B$  is true at  $x$ .

At an impossible world, logical truths—for example, of the form  $B \rightarrow B$ —may fail. This is achieved by giving the truth conditions of  $\rightarrow$  at such a world,  $w$ , in terms of a ternary relation,  $R$ :

$A \rightarrow B$  is true at  $w$  iff for all worlds  $x, y$ , such that  $Rwx$ , when  $A$  is true at  $x$ ,  $B$  is true at  $y$ .

These semantics give the base member of the family of logics,  $B$ . Other logics in the same family may be obtained by adding constraints on the relation  $R$ . The Anderson/Belnap logic,  $R$ , is one requiring a number of such constraints. At the time of writing, the nature of  $R$ , and so of plausible constraints on it, are still contentious issues.

Another important feature of the semantics of relevant logics is their handling of negation. If  $(A \& \neg A) \rightarrow B$  is not to be a logical truth, there must be worlds at which  $A \& \neg A$  holds (bringing out the connection between relevant logic and paraconsistent logic). This may be achieved in (at least) two ways. In the first (due originally to Dunn), formulas may take the values *true* and *false* independently (and so may take both or neither). The truth/falsity conditions for negation at a world,  $w$ , are then:

$\neg A$  is true at  $w$  iff  $A$  is false at  $w$

$\neg A$  is false at  $w$  iff  $A$  is true at  $w$

If  $A$  is both true and false at  $w$ , so is  $\neg A$ . So (given the natural semantics for  $\&$ )  $A \& \neg A$  is true (and false) at  $w$ .

The second way to handle negation is to treat truth and falsity as usual, but to use the “Routley  $*$ ”—invented by Valerie Routley (later Plumwood) and Sylvan. For each world,  $w$ , there is a world  $w^*$  (usually taken to satisfy the condition that  $w = w^{**}$ .) The truth conditions for negation are:

$\neg A$  is true at  $w$  iff  $A$  is false at  $w^*$

If  $w$  is  $w^*$ , then exactly one of  $A$  and  $\neg A$  holds at  $w$ . But if  $w$  is distinct from  $w^*$ , and  $A$  is true at  $w$  and false at  $w^*$ , then  $A \& \neg A$  is true at  $w$ . Again, at the time of writing, the philosophical meaning of  $*$  is still a contentious issue.

However the semantics of negation is handled, there will be worlds where  $A$  and  $\neg A$  hold; and so, assuming the standard behavior of disjunction, where  $\neg A \vee B$  holds, for arbitrary  $B$ . It follows that the disjunctive syllogism ( $A, \neg A \vee B \vdash B$ ) is invalid. This is significant because it shows that the ramifications of relevant logic spread much wider than may have been thought. In particular, the syllogism does not seem inherently dubious in the same way that the paradoxes of the material conditional are. The invalidity of the syllogism has therefore occasioned much of the criticism attracted by relevant logic. Defenders of relevant logic have replied in various ways.

Philosophical critiques aside, relevant logics have turned out to have a number of interesting mathematical properties. For example,  $R$  and some of the other stronger logics (though not the weaker ones) have the unusual property (for a propositional logic) of being undecidable (as shown by Urquhart). Relevant logics are intimately related with algebraic structures called De Morgan lattices, and can also be shown to fit in to the more general class of substructural logics.

**See also** Logic, Non-Classical; Modal Logic; Paraconsistent Logics.

### Bibliography

- Anderson, Alan, and Nuel Belnap. *Entailment: The Logic of Relevance and Necessity*. Vol. 1. Princeton, NJ: Princeton University Press, 1975.
- Anderson, Alan, Nuel Belnap, and J. Michael Dunn. *Entailment: The Logic of Relevance and Necessity*. Vol. 2. Princeton, NJ: Princeton University Press, 1992.
- Dunn, J. Michael. "Relevance Logic and Entailment." In *Handbook of Philosophical Logic*. Vol 3., 117–224, edited by Dov M. Gabbay and Franz Guenther. Dordrecht, Holland: Kluwer, 1986. The article is revised, with Greg Restall as a co-author, as "Relevant Logic" in the second edition of the *Handbook*. Vol. 6, 1–128. Dordrecht, Holland: Kluwer, 2002.
- Mares, Edwin. *Relevant Logic: A Philosophical Interpretation*. Cambridge, U.K.: Cambridge University Press, 2004.
- Mares, Edwin D., and Robert K. Meyer. "Relevant Logic." In *The Blackwell Guide to Philosophical Logic*, edited by L. Goble, 280–308. Oxford: Blackwells, 2001.
- Priest, Graham. *Introduction to Non-Classical Logic*. Chaps. 8–10. Cambridge, U.K.: Cambridge University Press, 2001.
- Read, Stephen. *Relevant Logic: A Philosophical Examination of Inference*. Oxford: Blackwells, 1988.
- Restall, Greg. *An Introduction to Sub-Structural Logics*. London: Routledge, 2000.
- Routley, Richard, Val Plumwood, Robert K. Meyer, and Ross Brady. *Relevant Logics and Their Rivals*. Vol 1. Atascadero, CA: Ridgeview, 1982. Vol. 2, edited by Ross Brady, Aldershot, U.K.: Ashgate, 2003.

**Graham Priest (2005)**

## RELEVANT ALTERNATIVES

To know a proposition, is it necessary that one is able to rule out every possibility of error associated with that proposition? Notoriously, infallibilism about knowledge—as defended, for example, in early work by Peter Unger (1975)—demands just this and argues on this basis for the skeptical conclusion that knowledge is rarely, if ever, possessed. Intuitively, however, the answer to this question is “no,” in that in everyday life we only demand that knowers rule out those error-possibilities that are in some sense relevant. For example, to know that the bird before me is a goldfinch, I may be required to be able to rule out that it is not some other bird that could be in the area just now, like a jackdaw, but we would not normally demand (at least not without special reasons) that I be able to rule out the possibility that it is not a mechanical goldfinch made up to be an exact replica of the real thing.

If this line of thought is right, then this prompts a relevant alternatives (RA) theory of knowledge that demands that one only needs to be able to rule out all relevant error-possibilities in order to know, not that one is able to rule out all error-possibilities, even irrelevant ones. (A similar view could be applied to other epistemic notions, like warrant or justification. For simplicity, the focus here is on knowledge.) Such a position would thus be a form of fallibilism, which is directly opposed to infallibilism and which thereby counters those versions of skepticism that are based on infallibilist considerations. The task at hand for the RA theorist is to offer a principled account of what makes an alternative relevant.

### RELEVANT ALTERNATIVES AND SENSITIVITY

One can find the beginnings of an RA theory of knowledge in the writings of such figures as Ludwig Wittgenstein and John Austin. The first worked out versions of an RA theory, however, can be found in the works of Fred Dretske (1970) and Robert Nozick (1981), who primarily understand knowledge in terms of the possession of beliefs that are sensitive to the truth in the following manner:

#### *Sensitivity*

An agent,  $S$ , has a sensitive belief in a true contingent proposition,  $p$ , if and only if, in the nearest possible worlds in which  $p$  is not true,  $S$  no longer believes  $p$ .

To illustrate this, consider again the example of the goldfinch discussed earlier. Given that the actual world is

roughly as we take it to be, gaining a sensitive belief in the proposition, P, that there is a goldfinch before one is relatively straightforward. All one needs is a true belief in this regard and, in the nearest possible worlds where P is no longer true—where, for example, the goldfinch has flown away leaving behind just an empty branch—one no longer believes that there is a goldfinch there, as presumably one does not. Notice that the relevant possible worlds here are limited and concern error-possibilities (e.g., that there is nothing at all on the branch rather than a goldfinch), which are easy to rule out. A theory of knowledge that treats sensitivity as the key requirement on the acquisition of knowledge is thus in a good position to capture the intuition that ordinarily we do not demand that agents are able to eliminate all possibilities of error before we count them as possessing knowledge.

Interestingly, however, the sensitivity-based approach does treat far-off possible worlds, and thus far-fetched error-possibilities, as sometimes being relevant to the possession of knowledge. Consider, for example, the hypothesis Q, that there is a mechanical goldfinch before one, constructed in such a way as to be indistinguishable to the naked eye from the real thing. When one is faced with what seems to be a goldfinch (and circumstances are, apparently, entirely normal), does one know not-Q? According to the sensitivity-based account of knowledge, this is unlikely because it is difficult to have a sensitive belief in not-Q. After all, to have a sensitive belief in this proposition it would be necessary to have a belief that was not only true in the actual world, but that was also no longer held in the nearest possible worlds in which not-Q is false—that is, those worlds in which Q is true, where one is at present looking at a mechanical goldfinch. The problem is, of course, that, *ex hypothesi*, one would continue to believe that one is looking at a real goldfinch even when one is faced with a mechanical goldfinch, at least unless one conducted special tests (such as capturing the “creature” and cutting it open). So while knowing P is relatively easy, knowing not-Q is hard. And notice that the reason this is the case is because the range of possible worlds, and thus the range of error-possibilities, that is relevant to the determination of one’s knowledge is different in each case.

#### RELEVANT ALTERNATIVES AND NONCLOSURE

On the face of it, this rendering of the RA theory seems to capture our pretheoretical intuition that in normal circumstances we ought to be able to know that we are looking at a goldfinch even though we are unable to rule out

(i.e., know to be false) the hypothesis that we are looking at a mechanical goldfinch. Nevertheless, this view does have a counterintuitive result, one that both Dretske and Nozick are prepared to accept. This is that the highly intuitive principle that knowledge is “closed” under known entailment (“closure”) has to be rejected. We can roughly formulate closure as follows:

##### *Closure for Knowledge*

If an agent, S, knows a proposition, p, and S knows that p entails a second proposition, q, then S also knows q.

For example, if one knows P, then given that one also knows that P entails not-Q (as surely one does), it follows from closure that one must know not-Q. Conversely, of course, if one fails to know the latter proposition, which is what the sensitivity-based approach predicts, then one fails to know the former.

Closure is highly intuitive and yet, as we have just seen, if it holds it would appear to license a restricted form of infallibilism. For although closure does not demand that it is a precondition on knowledge possession that one is able to rule out all possibilities of error, it does demand that one is able to rule out (i.e., know to be false) all those error-possibilities that are known to be inconsistent with what one knows, and this set of error-possibilities, while smaller, is large enough. This point is important, since if the appeal of infallibilism rests on the appeal of closure, then the view is on far stronger ground that one might have initially supposed because of the obvious appeal of the closure principle.

Nevertheless, Dretske and Nozick argue that recognizing that sensitivity is a necessary condition for knowledge highlights why this principle must go, since there are clearly cases, such as the goldfinch example, where one knows one proposition (and thus has a sensitive belief in this proposition) and knows that this proposition entails a second proposition, and yet one lacks a sensitive belief in the entailed proposition and thus fails to know it.

#### RELEVANT ALTERNATIVES AND CONTEXTUALISM

Although the sensitivity-based proposal has been influential, it does face the problem that it denies the highly intuitive closure principle for knowledge, and this has led some commentators to try to see if there is a way of accommodating the general intuition behind the RA theory in a way that preserves this principle. One of the guiding considerations behind views that try to offer an RA thesis that is consistent with closure is that the Dretske-

Nozick treatment seems to incorporate the idea that closure fails because while sometimes knowledge is hard to attain, sometimes attaining it is relatively straightforward. This tends to suggest that an alternative way of approaching the issue could be to regard knowledge as in some sense context-sensitive, so that one knows both of the target propositions in the closure-based inference relative to one set of epistemic standards (the less demanding ones), but knows neither of them relative to another set of epistemic standards (the more demanding ones). We would thus get a view that incorporates a reading of the RA intuition—because it would remain that not every error-possibility is always relevant to the possession of knowledge—but which was also consistent with closure. This view—known as contextualism about knowledge—is hinted at in an early response to Dretske’s denial of closure written by Gail Stine (1976), and has been developed by Stewart Cohen (1991), Keith DeRose (1995), and David Lewis (1996).

Consider again the goldfinch example. On the Dretske-Nozick view the class of possible worlds, and thus the class of error-possibilities, that is relevant to the determination of knowledge can differ depending on the content of the proposition at issue, which is why in this case the agent comes out as knowing P while failing to know not-Q, despite knowing that the former entails the latter. The reason for this is that when it comes to knowing P only nearby possible worlds are relevant, whereas knowing not-Q brings in farther out possible worlds. Suppose instead, however, that one simply treated the class of possible worlds as fixed in each context, so that the epistemic status of all beliefs—whatever their content—were in that context evaluated relative to those possible worlds. In normal contexts, then, only nearby possible worlds would be relevant, while in more demanding contexts far-off possible worlds would become relevant. This way of understanding knowledge would mean dropping sensitivity as a requirement on knowledge, of course, since there may be no nearby possible worlds in which the target proposition is false (this is, indeed, what we would expect to be the case when it comes to the hypothesis that one is at present looking at a mechanical goldfinch). Nevertheless, the guiding thought here is that so long as the agent’s belief matches the truth in the relevant possible worlds—that is, where the agent believes that proposition, it is true; and where the proposition is not true, the agent does not believe it—then the agent’s belief will be in the market to be counted as an instance of knowledge.

By contextualist lights, then, in contexts where the epistemic standards are low (and thus only nearby possible worlds count as relevant) one will tend to know both P and not-Q, since even one’s belief in not-Q will tend to match the truth (i.e., one believes it in all nearby possible worlds and it is true in all nearby possible worlds). In contrast, in contexts where the epistemic standards are more demanding, and thus where farther out possible worlds become relevant, it will now no longer be the case that one will tend to know either of these propositions. After all, there will be possible worlds, such as the far-off world in which there is a sophisticated plot to deceive people about the presence of goldfinches, in which one’s beliefs in P and not-Q no longer match the truth. Thus, as long as one consistently sticks to a specific epistemic standard then this construal of the RA intuition is not in conflict with closure since, depending on the context at issue, either one has knowledge of both of the target propositions or one has knowledge of neither of them.

#### RELEVANT ALTERNATIVES AND SAFETY

In more recent work, however, a third rendering of the RA thesis has come to the fore, one that is neither contextualist nor results in the denial of closure. This position—defended, for example, by Ernest Sosa (1999)—holds that far-off possible worlds are always irrelevant to knowledge, whatever the content of the target proposition or the context at issue. Accordingly, one is able to know, for example, both P and not-Q, whatever the context.

This view tends to hold that the key condition that a belief must meet if it is to count as knowledge is that it be safe. Safety can be roughly formulated as follows:

##### *Safety*

An agent, S, has a safe belief in a true contingent proposition, p, if and only if, in all nearby possible worlds in which S believes p, p is true.

Notice that contextualists will have to appeal to something like safety to explain how agents can know a proposition like not-Q in epistemically undemanding contexts where there are no nearby worlds in which what is believed is false. The point will be that while such beliefs are not sensitive, since there is no relevant Q-world for them to be sensitive to, they are safe, in that the agent’s belief in not-Q is always true across the relevant possible worlds. What is important about safety for our purposes is that it simply specifies the class of possible worlds that is relevant and leaves the matter at that—there is no room

here for a shift in context that would in turn alter the class of possible worlds, and thus the class of error-possibilities, that is relevant to the determination of knowledge. Accordingly, it will not matter which context one is in. Just so long as one's beliefs that P and not-Q are both safe—as presumably they will be—then one is in a position to know both of these propositions and thus there is no tension with closure.

### CONCLUDING REMARKS

There are thus three competing conceptions of the RA intuition in the literature. The first view treats relevance as being determined by the content of the proposition known, and as a result maintains that the closure principle for knowledge fails. The second view treats relevance as being determined by context, and thereby retains closure. Finally, the third view also retains closure, but does so by maintaining an invariant standard of relevance, regardless of the content of the target proposition or of the context at issue.

*See also* Austin, John; Contextualism; Dretske, Fred; Lewis, David; Nozick, Robert; Propositions; Sosa, Ernest; Wittgenstein, Ludwig Josef Johann.

### Bibliography

- Austin, J. L. "Other Minds." In *Philosophical Papers*. Oxford, U.K.: Clarendon Press, 1961.
- Cohen, Stewart. "Skepticism, Relevance, and Relativity." In *Dretske and His Critics*, edited by Brian P. McLaughlin. Cambridge, MA: Basil Blackwell, 1991.
- DeRose, K. "Solving the Sceptical Problem." *Philosophical Review* 104 (1995): 1–52.
- Dretske, Fred. "Epistemic Operators." *Journal of Philosophy* 67 (1970): 1007–1023.
- Lewis, D. "Elusive Knowledge." *Australian Journal of Philosophy* 74 (1996): 549–567.
- Nozick, Robert. *Philosophical Explanations*. Cambridge, MA: Harvard University Press, 1981.
- Sosa, Ernest. "How to Defeat Opposition to Moore." *Philosophical Perspectives* 13 (1999): 141–154.
- Stine, Gail. "Skepticism, Relevant Alternatives, and Deductive Closure." *Philosophical Studies* 29 (1976): 249–261.
- Unger, Peter. *Ignorance: A Case for Scepticism*. Oxford, U.K.: Clarendon Press, 1975.
- Wittgenstein, Ludwig. *On Certainty*. Translated by Denis Paul and G. E. M. Anscombe, edited by G. E. M. Anscombe and G. H. von Wright. Oxford, U.K.: Blackwell, 1969.

**Duncan Pritchard (2005)**

## RELIABILISM

Reliabilism is an approach to the analysis of either knowledge or justified belief that makes, in some way or another, the reliability of belief-producing faculties or processes the key notion of epistemic assessment. An early version of a reliabilist theory of *knowledge* was proposed by David M. Armstrong (1973), who thought of knowledge in terms of a reliable thermometer that accurately indicates the correct temperature. A (noninferential) true belief amounts to knowledge, according to Armstrong, if its properties nomically (i.e., via the laws of nature) guarantee its truth. Closely related theories conceive of knowledge as resulting from a counterfactual guarantee of truth. For instance, according to Robert Nozick (1981), knowledge comes about when a subject's belief that p tracks the truth of p, which it does (focusing just on the core of Nozick's theory) if the following condition is met: S would not believe that p if p were false. Alternatively, Fred Dretske (1971, 1981) suggests that a true belief counts as knowledge if the subject possesses a conclusive reason for p. According to this proposal, S knows that p on the basis of a reason, R, if R would not be true unless p were true.

### RELIABILIST THEORIES OF JUSTIFIED BELIEF

Turning to reliabilist theories of *justified belief*, there are two main versions to consider: reliable indicator theories, and process reliabilism. A version of the former was developed by Marshall Swain (1979): What makes S's belief that p justified is the belief's being a reliable indication (conceived of in terms of objective probability) of p's truth. In contrast, process reliabilism (reliabilism<sub>pr</sub>), as advocated by Alvin Goldman (1979), focuses not on the belief itself, but its causal history, or the cognitive process by which it was formed. The basic proposal is that a belief that p is justified if and only if the process by which it was formed is (sufficiently) reliable. On the one hand, perception, memory, and introspection are reliable cognitive processes, typically resulting in justified beliefs and indeed knowledge. Hasty generalization, wild hunches, and wishful thinking, on the other hand, are unreliable processes and invariably produce unjustified beliefs and prevent the formation of knowledge. Moving beyond this initial characterization, an account of reliabilism<sub>pr</sub> requires refinement as to the question of precisely what determines the reliability of a cognitive process. According to one approach, a process's reliability is fixed by the truth ratio of its actual doxastic output: The greater the ratio of true over false beliefs within the set of beliefs that

make up a process's actual track record, the higher its degree of reliability. As an alternative to such a *track record* conception, William Alston (1995) recommends a *propensity* construal, according to which a process's reliability is determined by, not its actual track record, but what the truth ratio of its output *would be* in an appropriate range of cases.

According to reliabilism<sub>pr</sub>, as well as other versions of reliabilism, whether S's belief that p is justified, or an instance of knowledge, does not depend—at least not exclusively—on S's evidence relative to p. Reliabilist theories of knowledge and justification must thus be viewed as intended alternatives to evidentialist theories (Chisholm 1989, Conee and Feldman 1985). Nevertheless, in Goldman's defense of reliabilism<sub>pr</sub>, evidentialist considerations do not completely drop out of the picture. Suppose the following: (i) S forms the belief that p via process C; (ii) C is in fact reliable; and (iii) on the basis of professional testimony, S has reason to believe that C is unreliable. Intuitively, S's belief is not justified. Cases such as that make it doubtful that origination in a reliable process is by itself sufficient for justification. Accordingly, Goldman (1986) supplements his account with a nonundermining clause to the effect that S must not believe, or be in possession of evidence supporting the belief, that the relevant process is unreliable.

In its most radical and challenging manifestation, reliabilism<sub>pr</sub> asserts that reliable belief formation is both necessary and sufficient for a belief's justification. Alston (1989) defends a moderate version, according to which reliable belief formation is merely necessary for justification. An approach that may be viewed as an alternative to reliabilism<sub>pr</sub>—virtue epistemology—shifts the focus away from reliable processes to reliable faculties or cognitive virtues, giving rise to the thought that justification and knowledge may be conceived of as resulting from the employment of virtuous faculties (Sosa 1991). A related approach is advocated by Alvin Plantinga (1993). Plantinga's view is that knowledge is generated by properly functioning faculties, where a faculty's proper functioning requires reliability, in addition to adequate design and an orientation towards truth and the avoidance of falsehood.

### THE THREE MAJOR PROBLEMS WITH RELIABILISM

Reliabilism<sub>pr</sub> is confronted with three major problems. The first of these raises the issue of whether production via reliable processes is necessary for justification (Cohen 1984, Ginet 1985). Consider the beliefs of a subject who

is deceived by an evil demon. Because the evidential situation of an evil demon victim is not relevantly different from that of a normal person who has (presumably) by and large justified beliefs, it is commonly agreed that such a victim's beliefs are, just like a normal person's, by and large justified. Alas, the victim's beliefs are, unlike those of a normal person, massively false. The challenge for reliabilism<sub>pr</sub>, then, is this: The beliefs of an evil demon victim are justified although they are the result of *unreliable* cognitive processes.

There are four ways in which reliabilists can respond. First, they can—implausibly—deny that evil demon victims have justified beliefs. Second, they can deny the relevance of the counterexample, as Alston (1995) does. He argues that the cognitive processes to consider are to be restricted to those that would yield a high truth ratio over a wide range of situations *of the kind one typically encounters*. Third, advocates of reliabilism<sub>pr</sub> can try to accommodate the counterexample by modifying reliabilism<sub>pr</sub>. For example, Goldman (1986) introduces normal worlds reliabilism, the basic idea of which is that a belief (in any possible world) is justified if and only if the process by which it was formed is reliable in normal worlds: worlds that correspond to our general beliefs about the actual world. Because perception is reliable in normal worlds, normal worlds reliabilism arguably yields the result that the perceptual beliefs of evil demon victims are justified.

Whether this response succeeds depends on whether the processes by which an evil demon victim forms beliefs can properly be characterized as perceptual, memorial, and so on. After all, at the end of the causal chains from which a deceived subject's beliefs originate, there is the evil demon: not exactly the kind of creature one finds in any normal worlds. So if the belief-generating processes are considered in their entirety, it is hard to tell what their truth ratios in normal worlds would be because such processes are not part of any normal worlds to begin with. Fourth, defenders of reliabilism<sub>pr</sub> can introduce different concepts of justification. For example, Goldman (1988) distinguishes between strong and weak justification. According to this proposal, the beliefs of evil demon victims are justified in the weak sense, whereas reliabilism is intended to be an analysis of strong justification, the kind of justification that is needed if a true belief is to count as an instance of knowledge.

**THE SECOND PROBLEM.** The second problem raises the issue of whether origination in a reliable process is sufficient for justification (BonJour 1985, ch. 4). Suppose the following: (i) Norman's belief that p is the result of clair-

voyance; (ii) Norman's faculty of clairvoyance is reliable; and (iii) Norman has no reason to believe that his belief that *p* originated in, or is sustained by, a reliable faculty. Reliabilism<sub>pr</sub> implies that Norman is justified in believing that *p*, whereas intuitively he seems to be unjustified. In response, reliabilists can deny the intuition underlying the objection—that is, insist that Norman's belief is justified. Alternatively, they can, once again, attempt to accommodate the example by devising a suitable modification. For example, they might consider *ignoring the absence of evidential support* an unreliable cognitive process, and suggest that beliefs whose causal origin includes that process are unjustified. The problem with this suggestion is that it threatens to rob reliabilism<sub>pr</sub> of its identity by letting it collapse into a disguised version of evidentialism.

**THE THIRD PROBLEM.** The third problem, known as the “generality problem,” raises the issue of how to individuate the cognitive processes the reliability of which is supposed to determine whether a belief is, or fails to be, justified (Feldman 1985). Suppose a person sees, and thus believes, that the cat is lying on the couch. The process by which this belief is formed could plausibly be classified as perception. More specifically, it could be viewed as an instance of visual perception. Further specification yields further choices: visual perception at a distance of (say) eight feet; visual perception of a medium sized object at a distance of eight feet; visual perception of a medium sized object at a distance of eight feet under daylight illumination. In general terms, the point is that a particular *token* of a cognitive process instantiates many different process *types*. Some of them are reliable, some of them are not. Whether a belief whose justificational status is at issue comes out justified or unjustified will depend on which process type is made the basis of the assessment. The challenge advocates of reliabilism<sub>pr</sub> face is to give a principled account of how to select the right process type.

Consider perception, an obviously reliable process. But not all perceptual beliefs are justified. Nor are, for that matter, all memorial beliefs, or all visual beliefs, or all auditory beliefs. Sometimes, perceivers fail to take into account undermining evidence, and then beliefs produced by reliable processes fail to be justified. Hence individuating process types using broad categories such as perception, vision, or memory will often yield the wrong results. More specification is clearly required. But too much specification also yields the wrong results. At the extreme end of specification are process types instantiated by one and only one process token; one then

encounters what Richard Feldman (1985) calls the “single case problem.” If such a token results in a false belief, the result will be total unreliability, for the process type's output is false in all cases. If the process token in question results in a true belief, the result will be perfect reliability, for the process type's output is true in all cases. In the former case, the belief will be unjustified no matter what; in the latter case, it will be justified no matter what. This will result in clearly counterintuitive results. Suppose *S* is a paranoid schizophrenic. While riding on the bus, *S*'s paranoia leads him to believe that the bus will blow up. Suppose further that that is in fact true. Let *P\** stand for a process type described in such a way that *P\** has *one and only one* instantiation: the process token that caused *S*'s belief about the bus. (Such a description can easily be achieved by making reference to properties that uniquely pick out *S* and the circumstances under which *S* formed that belief.) Because *P\** is a (perfectly) reliable process, reliabilism<sub>pr</sub> implies—implausibly—that *S* is justified in believing that the bus will blow up. The problem for advocates of reliabilism<sub>pr</sub> is this: On the basis of what principled grounds can they claim that *P\** is not the process type the reliability of which determines the justificational status of *S*'s belief?

Alston (1995) claims that there is a solution to the generality problem. Regarding the single case problem, he suggests that it does not arise when the track-record conception of reliability is replaced with a propensity conception. Consider again the belief of the paranoid subject that the bus will blow up. On the propensity conception of generality, the process type in question, having precisely one instantiation that led in fact to a true belief, nevertheless counts as unreliable when it is taken into account that beliefs resulting from paranoia tend to be false. Moreover, Alston argues that there are objective, psychological facts of the matter that determine, for each process token leading to a particular belief, which process type this token instantiates. According to Alston, every process token instantiates an input-output function. Each time a belief is formed, Alston claims, there is one and only one input-output pair that is psychologically real. However, Earl Conee and Richard Feldman (1998) respond that, even if one accepts the constraints Alston places on the selection of legitimate process types, there will still be a wide range of process types going from narrow to broad characterizations. As a result, there will be cases of belief formation for which reliabilism<sub>pr</sub> will not yield a determinable implication about the belief's justificational status.



## INTERNALIST AND EXTERNALIST THEORIES

It is common practice to distinguish between *internalist* and *externalist* theories of knowledge and justification. According to internalism about justification, the factors that determine a belief's justificational status (call them "J-factors") must be internal to the subject's mind. Typically, such internality is defined epistemically: An item *x* is internal to *S* if and only if *S* can, merely by reflecting on it, determine whether *x* is present or absent. Internalist theories of justification, then, usually demand that J-factors must be such that their presence or absence is always on reflection recognizable by the subject. As a result of this constraint, justification itself turns into something the presence or absence of which can be recognized upon reflection. According to externalism about justification, J-factors are not subject to any internality constraint. Reliabilism about justification, in its various manifestations, is an externalist theory, for the obvious reason that subjects are not always in a position to determine, solely on the basis of reflection, whether their beliefs are the result of reliable cognitive processes. Consider, again, the victim of an evil demon. Upon reflection, such a victim will think that her perceptual beliefs originate in reliable cognitive processes, when in fact they do not. However, the classification of reliabilism as an externalist theory should not be misunderstood to mean that, according to reliabilism, the reliability of our cognitive processes is completely beyond our ken. To the contrary, there is no reason why reliabilists should deny that, in typical situations when a subject forms, for example, perceptual or memorial beliefs, it should be knowable to the subject, on the basis of presently available evidence, that the beliefs in question have their origin in reliable processes or faculties.

The internalism/externalism issue presents itself in a different form when the object of the dispute is not justification but knowledge. There is broad agreement that knowledge is not, and indeed cannot be, internal in the way in which, according to some, justification is internal. Suppose the following: (i) *S* has a body of excellent evidence, *E*, in support of *p*; (ii) *E* is misleading: *p* is in fact false. Reflecting on whether she knows that *p*, *S* will of course conclude, mistakenly, that she does. Clearly, then, whether or not one knows cannot always be determined upon reflection. Thus it is beyond dispute that knowledge is external. Nevertheless, it would not be inaccurate to say that evidentialists defend an internalist conception of knowledge. According to evidentialists, *S* knows that *p* only if *S* has a good reason for *p*. But whether or not one

has a good reason for *p* is something that is internal to the subject; it is something that can be determined merely by reflecting on one's evidence. Evidentialists, therefore, hold that one of the necessary conditions of knowledge is internal, and thus may be considered internalists about knowledge. Reliabilists, however, are externalists about knowledge, for they typically claim that reliable belief production, suitably qualified, is sufficient for making a true belief an instance of knowledge, thus advocating an account of knowledge without any internalist condition.

The reliabilist, externalist view that the employment of reliable processes or faculties is sufficient for making a true belief thus produced an instance of knowledge that can be supported by citing that very young children and animals possess knowledge, for neither the former nor the latter would seem to be capable of having good reasons in support of their beliefs. The evidentialist, internalist view that one knows only if one possesses a good reason can be defended by pointing out that, upon discovering that a person believes that *p* without having a good reason for *p*, we tend to judge that that person does not know that *p*. Reliabilists, on the one hand, need to come to terms with one kind of fact about our ordinary cognitive practice: people are reluctant to attribute knowledge in the absence of good reasons. Evidentialists, on the other hand, need to come to terms with another kind of fact about our ordinary cognitive practice: people do not hesitate to attribute knowledge to very young children and even such animals as cats and dogs.

## CONCLUSION

Even though reliabilist theories are properly classified as externalist, there is no reason in principle why internalists should not acknowledge the relevance, or even fundamental importance of reliability. To begin with, internalists might agree that a true belief counts as knowledge only if it originates in a reliable faculty. Furthermore, Matthias Steup (2004) proposes *internalist reliabilism* as an answer to the question of why sense experience is a source of justification. According to Steup's proposal, what makes sense experience a source of justification is not *de facto* reliability, but *evidence* of reliability. Perception is a source of justification for a subject, *S*, if and only if *S* has, on the basis of track record memories, reason to believe that her perceptual faculties are reliable. According to this proposal, perception is a source of justification even in worlds in which it is in fact unreliable, as long as, on the basis of adequate evidence, it appears to be reliable. According to such an approach, issues of reliability lie indeed at the heart of epistemology, for *S* acquires

knowledge only if S (i) employs a faculty that is in fact reliable, and (ii) possesses evidence of that faculty's reliability.

**See also** Alston, William P.; Armstrong, David M.; Dretske, Fred; Epistemology; Epistemology, History of; Evidentialism; Goldman, Alvin; Nozick, Robert; Plantinga, Alvin; Virtue Epistemology.

### **Bibliography**

- Alston, William. "How to Think about Reliability." *Philosophical Topics* 23 (1995): 1–29.
- Alston, William. "An Internalist Externalism." In *Epistemic Justification: Essays in the Theory of Knowledge*. Ithaca, NY: Cornell University Press, 1989.
- Armstrong, D. M. *Belief, Truth, and Knowledge*. Cambridge, U.K.: Cambridge University Press, 1973.
- BonJour, Laurence. *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press, 1985.
- Chisholm, Roderick. *The Theory of Knowledge*. Englewood Cliffs, NJ: Prentice Hall, 1989.
- Cohen, Stewart. "Justification and Truth." *Philosophical Studies* 46 (1984): 279–95.
- Conee, Earl, and Richard Feldman. "Evidentialism." *Philosophical Studies* 48 (1985): 15–44.
- Conee, Earl, and Richard Feldman. "The Generality Problem for Reliabilism." *Philosophical Studies* 89 (1998): 1–29.
- Dretske, Fred. "Conclusive Reasons." *Australasian Journal of Philosophy* 49 (1971): 1–22.
- Dretske, Fred. *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press, 1981.
- Feldman, Richard. "Reliability and Justification." *Monist* 68 (1985): 159–74.
- Ginet, Carl. "Contra Reliabilism." *Monist* 68 (1985): 175–87.
- Goldman, Alvin. "Epistemic Folkways and Scientific Epistemology." In *Liaisons: Philosophy Meets the Cognitive and Social Sciences*. Cambridge, MA: MIT Press, 1991.
- Goldman, Alvin. *Epistemology and Cognition*. Cambridge, MA: Harvard University Press, 1986.
- Goldman, Alvin. "Strong and Weak Justification." *Philosophical Perspectives* 2 (1988): 51–69.
- Goldman, Alvin. "What Is Justified Belief?" In *Justification and Knowledge*, edited by George Pappas. Dordrecht, Netherlands: Reidel, 1979.
- Nozick, Robert. *Philosophical Explanations*. Cambridge, MA: Harvard University Press, 1981.
- Plantinga, Alvin. *Warrant: The Current Debate*. Oxford: Oxford University Press, 1993.
- Sosa, Ernest. *Knowledge in Perspective*. Cambridge, U.K.: Cambridge University Press, 1991.
- Steup, Matthias. "Internalist Reliabilism." *Philosophical Issues* 14 (2004).
- Swain, Marshall. *Reasons and Knowledge*. Ithaca, NY: Cornell University Press, 1979.

**Matthias Steup (2005)**

## RELIGION

This entry is not a survey of the various forms that "religion" has taken in human history; rather, it treats the nature of religion as a problem in the philosophy of religion. It will be concerned with attempts to develop an adequate definition of religion, that is, to make explicit the basic features of the concept of religion.

### GENERAL DEFINITION AND CHARACTERISTICS

EXAMINATION OF DEFINITIONS. A survey of existing definitions reveals many different interpretations.

"Religion is the belief in an ever living God, that is, in a Divine Mind and Will ruling the Universe and holding moral relations with mankind."  
—James Martineau

"Religion is the recognition that all things are manifestations of a Power which transcends our knowledge."  
—Herbert Spencer

"By religion, then, I understand a propitiation or conciliation of powers superior to man which are believed to direct and control the course of Nature and of human life."  
—J. G. Frazer

"Religion is rather the attempt to express the complete reality of goodness through every aspect of our being."  
—F. H. Bradley

"Religion is ethics heightened, enkindled, lit up by feeling."  
—Matthew Arnold

"It seems to me that it [religion] may best be described as an emotion resting on a conviction of a harmony between ourselves and the universe at large."  
—J. M. E. McTaggart

"Religion is, in truth, that pure and reverential disposition or frame of mind which we call piety."  
—C. P. Tiele

"A man's religion is the expression of his ultimate attitude to the universe, the summed-up meaning and purport of his whole consciousness of things."  
—Edward Caird

"To be religious is to effect in some way and in some measure a vital adjustment (however tentative and incomplete) to whatever is reacted to or regarded implicitly or explicitly as worthy of serious and ulterior concern."  
—Vergilius Ferm

If we take these definitions as attempts to state necessary and sufficient conditions for something to be a religion, it is not difficult to show that none of them is adequate. With respect to necessary conditions, consider Martineau's definition. It is clear that such a belief does not have to be present in a religion. No polytheistic religion recognizes a single divine ruler of the universe; and there are religions, such as Hinayana Buddhism, in which beliefs in personal deities play no role at all. Bradley and Arnold identify religion with morality, but there are primitive societies in which there is no real connection between the ritual system, with its associated beliefs in supernatural beings, and the moral code. The latter is based solely on tribal precedent and is not thought of as either originating with or sanctioned by the gods. If, as would commonly be done, we call the former the religion of the culture, we have a religion without morality. As for McTaggart and Tiele, it seems likely that if we specify "piety" or "feeling of harmony" sufficiently to give them a clear and unambiguous meaning, we will be able to find acknowledged religions in which they do not play an important role. It would seem that we could avoid this only by construing "piety," for example, to cover any state of feeling that arises in connection with religious activities. It does seem plausible to regard some of the definitions as stating necessary conditions, as in Caird and Ferm. However, it is doubtful that these are sufficient conditions. Does any "ultimate attitude" or any "vital adjustment" constitute a religion? As William James points out (*The Varieties of Religious Experience*, Ch. 2), it seems doubtful that a frivolous attitude toward life constitutes a religion, even if it is the fundamental attitude of a given person. And Ferm's overcarefully worded statement would seem to admit any attitude with respect to anything considered important to the ranks of the religious. This would presumably include one's attitude toward one's spouse, toward one's vocation, and, in many cases, toward one's athletic activities. At this point one wonders what has happened to the concept of religion. Many of the definitions are deficient on grounds of both necessity and sufficiency. To return to Martineau, it is quite conceivable that such a belief might be held purely as a speculative hypothesis, without affecting the believer's feelings and attitudes in the way that would be requisite for religious belief. And as for McTaggart, it seems clear that one could from time to time have such a sense of harmony without this being integrated into anything that we would call a religion.

It is noteworthy that most of these definitions stress one aspect or another of religion to the exclusion of others. Thus, Martineau and Spencer represent religion as

some sort of belief or other cognitive state; Frazer, as ritual (conceived in a utilitarian fashion); Bradley and Arnold, as a kind of moral attitude and activity; and McTaggart and Tiele as a certain kind of feeling. One might attribute the failings of these definitions to their one-sidedness. One could hardly expect to get an adequate statement of the nature of so complex a phenomenon as religion, essentially involving, as it does, all these forms of human activity by restricting oneself to belief, feeling, ritual, or moral attitude alone. Caird and Ferm escape this particular failing by concentrating on a comprehensive term such as *attitude* or *adjustment*, which itself embraces belief, feeling, and moral attitude. But, as we have seen, these formulations do not come measurably closer to providing a set of necessary and sufficient conditions.

There are other ways of construing definitions of religion. Instead of taking the above statements as attempts to specify features that are common and peculiar to cases of religion, we might take each of them as an attempt to state the *essence* of religion, that central feature in terms of which all religious phenomena are to be understood. This approach to the matter is explicit in the following statements:

"The essence of religion is a belief in the persistency of value in the world."—Harald Høffding

"The heart of religion, the quest of the ages, is the outreach of man, the social animal, for the values of the satisfying life."—A. E. Haydon

"The essence of religion consists in the feeling of an absolute dependence."—Friedrich Schleiermacher

There are two distinguishable interpretations of claims of this type. They might be interpreted genetically, as accounts of the origin of religion. The claim would then be that what is specified as the essence of religion is the original root from which all phenomena of religion have sprung. Thus, Julian Huxley, like Schleiermacher working with a conception of the essence of religion as a kind of feeling, says, "the essence of religion springs from man's capacity for awe and reverence, that the objects of religion ... are in origin and essence those things, events, and ideas which arouse the feeling of sacredness" (*Religion without Revelation*, p. 111). Similarly starting with Høffding's formulation, we might try to show how typical religious doctrines, rites, and sentiments grew out of an original belief in the persistency of value. However, since we know virtually nothing about the prehistoric origins of religion, speculation in this area is almost completely unchecked by data, and it seems impossible to find

any rational basis for choosing between alternative genetic accounts.

However, we might also give a nongenetic interpretation. Saying that the essence of religion is a feeling of absolute dependence, for example, might mean that the full interrelatedness of the various features of religion can be understood only if we view them all in relation to a feeling of absolute dependence. This claim would be independent of any view of the origin of religion. The difficulty with this is that there would seem to be several different features of religion that could be taken as central—such as ritual, a need for reassurance against the terrors of life, or a need to get a satisfactory explanation of the cosmos—and it is illuminating to view the rest of religion as related to each of these. How is one to settle on a unique essence?

**CHARACTERISTIC FEATURES OF RELIGION.** Despite the fact that none of the definitions specifies a set of characteristics which is present when and only when we have a religion, or gives us a unique essence, it does seem that they contribute to our understanding of the nature of religion. It appears that the presence of any of the features stressed by these definitions will help to make something a religion. We might call such features, listed below, religion-making characteristics.

- (1) Belief in supernatural beings (gods).
- (2) A distinction between sacred and profane objects.
- (3) Ritual acts focused on sacred objects.
- (4) A moral code believed to be sanctioned by the gods.
- (5) Characteristically religious feelings (awe, sense of mystery, sense of guilt, adoration), which tend to be aroused in the presence of sacred objects and during the practice of ritual, and which are connected in idea with the gods.
- (6) Prayer and other forms of communication with gods.
- (7) A worldview, or a general picture of the world as a whole and the place of the individual therein. This picture contains some specification of an overall purpose or point of the world and an indication of how the individual fits into it.
- (8) A more or less total organization of one's life based on the worldview.
- (9) A social group bound together by the above.

*Interrelations of characteristics.* Religion-making characteristics do not just happen to be associated in reli-

gion; they are intimately interconnected in several ways. Some of these connections have been indicated, but there are others. For example, the distinction between sacred and profane objects is based on other factors mentioned. It is not any intrinsic characteristic of a thing that makes it a sacred object; things of every conceivable kind have occupied this position—animals, plants, mountains, rivers, persons, and heavenly bodies. Certain objects are singled out as sacred in a given community because they typically arouse such feelings as awe and a sense of mystery, and thus the members of that community tend to respond to these objects with ritual acts. Again, the emotional reaction to sacred objects may be rationalized by conceiving the object to be the habitation or manifestation of a god. The awe aroused by the wild bull led to its being identified with the wild god of intoxication, Dionysus. The very special impression made by Jesus of Nazareth on certain of his contemporaries was expressed by calling him the Son of God. These examples make it sound as if emotional reactions to sacred objects come first and that these reactions are then explained by positing gods as their causes. But it can also happen the other way round. The acceptance of beliefs about the gods and their earthly habitations can contribute to the evocation of awe and other feelings in the presence of certain objects. The members of a religious community are taught to hold certain objects in awe by being taught various doctrines about the gods. Thus, Christians are taught to regard the cross and the consecrated bread and wine with reverence by being told of the Crucifixion and the Last Supper.

A similar reciprocal relationship holds between ritual and doctrine. A doctrine can be introduced as the justification of an already established ritual. Thus, the myth of Proserpine being carried off to the underworld and remaining there half the year seems to have been introduced as an explanation of a preexisting magical fertility cult, in which an ear of grain, perhaps called the corn maiden, was buried in the fall and raised sprouting in the spring. On the other hand, changes in doctrine can engender, modify, or abolish rituals. Beliefs about the divine status of Jesus Christ played an important role in shaping the Christmas festival.

*Definition in terms of characteristics.* If it is true that the religion-making characteristics neither singly nor in combination constitute tight necessary and sufficient conditions for something being a religion, and yet that each of them contributes to making something a religion, then it must be that they are related in some looser way to the application of the term. Perhaps the best way to put it

is this. When enough of these characteristics are present to a sufficient degree, we have a religion. It seems that, given the actual use of the term *religion*, this is as precise as we can be. If we tried to say something like “for a religion to exist, there must be the first two plus any three others,” or “for a religion to exist, any four of these characteristics must be present,” we would be introducing a degree of precision not to be found in the concept of religion actually in use.

Another way of putting the matter is this. There are cultural phenomena that embody all of these characteristics to a marked degree. They are the ideally clear paradigm cases of religion, such as Roman Catholicism, Orthodox Judaism, and Orphism. These are the cases to which the term *religion* applies most certainly and unmistakably. However, there can be a variety of cases that differ from the paradigm in different ways and to different degrees, by one or another of the religion-making characteristics dropping out more or less. For example, ritual can be sharply de-emphasized, and with it the demarcation of certain objects as sacred, as in Protestantism; it can even disappear altogether, as with the Quakers. Beliefs in supernatural beings can be whittled away to nothing, as in certain forms of Unitarianism, or may never be present, as in certain forms of Buddhism. And, as mentioned earlier, in certain primitive societies morality has no close connection with the cultic system. As more of the religion-making characteristics drop out, either partially or completely, we feel less secure about applying the term *religion*, and there will be less unanimity in the language community with respect to the application of the term. However, there do not seem to be points along these various dimensions of deviations that serve as a sharp demarcation of religion from nonreligion. It is simply that we encounter less and less obvious cases of religion as we move from, for example, Roman Catholicism through Unitarianism, humanism, and Hinayana Buddhism to communism. Thus, the best way to explain the concept of religion is to elaborate in detail the relevant features of an ideally clear case of religion and then indicate the respects in which less clear cases can differ from this, without hoping to find any sharp line dividing religion from nonreligion. (Cf. Ludwig Wittgenstein’s notion of “family-resemblances” among the things to which a term applies.)

An adequate definition of religion should throw light on the sorts of disputes and perplexities that typically produce a need to define religion, such as disputes over whether communism is a religion, and whether devotion to science can be called a man’s religion. So long as we are

dealing with definitions of the simplistic type that we have criticized, these problems are not illuminated. Each party to the dispute will appeal to a definition suited to the position he is defending, and since none of these definitions is wholly adequate, there is an irreducible plurality of not wholly inadequate definitions to be used for this purpose. Person *A*, who claims that communism is a religion, will give, for instance, Caird’s statement as his definition of religion, and person *B*, who denies this, will choose Martineau’s. Obviously, the position of each is upheld by his chosen definition. Hence, it would seem that the only way to settle the dispute is to determine which is the correct definition. However, we have seen that this gets us nowhere; no such definition is wholly adequate.

At this point there is a temptation to brand the dispute purely verbal, a reflection of different senses attached to the word *religion*. It may seem that the disagreement can be dissolved by persuading all parties to use the word in the same sense. But this is a superficial reaction that does not adequately bring out how much the parties to the dispute have in common. In fact, Martineau and Caird represent two contrasting emphases within a common framework. Suppose that *A* and *B* begin with the same paradigm, orthodox Protestant Christianity. But *A* gives greatest weight to the moral-orientation–emotion elements in this paradigm. As long as anything strongly manifests these elements, as long as it serves as a system of life orientation for the individual who is bound to it by strong emotional ties, he will call it a religion. *B*, on the other hand, gives greatest weight to the belief in a personal God and the complex of emotions, ritual, and devotional acts that is bound up with that belief. Thus, although they have basically the same concept of religion, they will diverge in their application of the term at certain points. Once we realize that this is the true situation, we can state the problem in a more tractable form. We can enumerate the religion-making characteristics and determine which of them communism has and in what degree. Then we can proceed to the heart of the dispute—the relative importance of these characteristics. Insofar as there is a real issue between *A* and *B*, once both are in possession of all the relevant facts, it is whether communism is similar to clear cases of religion in the most important respects, that is, whether the respects in which it is like Protestant Christianity are more important than those in which it is different.

## TYPES OF RELIGION

In the case of so complex a concept as religion, it is desirable to supplement the very general portrayal of basic fea-

tures with some indications of the varying emphases placed on them in different religions. To do this, we must develop a classificatory scheme.

William James has reminded us that in every religion there is some sort of awareness of what is called divine and some sort of response to this divinity. This being the case, a very fruitful way of classifying religions is to ask in the case of each: "Where is the divine (the object of religious responses) primarily sought and located, and what sort of response is primarily made to it?" In answering these questions for a given religion, the religion-making features most stressed in that religion will also come to light. According to this principle of division, religions fall into three major groups: sacramental, prophetic, and mystical.

**LOCATION OF THE DIVINE.** In sacramental religion the divine is sought chiefly in things—inanimate physical things like pieces of wood (relics of saints, statues, crosses), food and drink (bread and wine, baptismal water), living things (the totem animal of the group, the sacred cow, the sacred tree), processes (the movements of the sacred dance). This does not mean that the thing itself is responded to as divine, although this can happen in very primitive forms of sacramental religion, called fetishism. Usually the sacred thing is conceived to be the habitation or manifestation of some god or spirit. Thus, the ancient Hebrews treated the elaborate box that they called the Ark of God as the habitation of their god, Yahweh; the Hindus consider the river Ganges sacred to the god Shiva—they believe that Shiva is in some specially intimate relation to that river, and they bathe in its waters to benefit from his healing power. The Roman Catholic finds the presence of God concentrated in the consecrated bread and wine, which, he believes, has been transformed into the body and blood of Christ. At a more sophisticated level the material thing may be taken as a symbol of the divine rather than as its direct embodiment, as in the definition of a sacrament given in the Anglican Book of Common Prayer, "an outward and visible sign of an inward and spiritual grace."

In prophetic religion the divine is thought to manifest itself primarily in human society—in the events of human history and in the inspired utterances of great historical figures. It is not denied that nature issues from the divine and is under divine control, but it is not in nature that God is most immediately encountered. The divine reality is to be discovered in great historical events—the destruction of cities, the rise and fall of empires, the escape of a people from bondage. The hand of God is

seen in these matters because God is encountered more immediately in the lives and the inspired words of his messengers, the prophets, who reveal in their utterances God's nature, his purposes and commands, and derivatively in the sacred books that contain the records of these revelations. Christianity, Judaism, and Islam, the three chief prophetic religions, are sometimes called religions of the book. Here the key term is not *sacrament* but *revelation*. Prophetic religion, unlike the others, stresses the word as the medium of contact with the divine. (An example is the opening of the Gospel of John.) For the ritualist, and still more for the mystic, whatever words he may use, the consummation of his endeavors is found in a wordless communion with the divine. In prophetic religion, however, the linguistic barrier is never let down; it is not felt as a barrier at all.

The center of mystical religion is the mystical experience, which at its highest development dominates the consciousness, excluding all awareness of words, nature, even of the mystic's own self. In this experience the individual feels himself pervaded and transformed by the divine, identified with it in an indivisible unity. The world and all its ordinary concerns seem as naught as the mystic is caught up in the ineffable bliss of this union. It is not surprising that those who have enjoyed this experience, and those who aspire to it, should take it to be the one true avenue of contact with the divine and dismiss all other modes as spurious, or at least as grossly inferior. Rituals and sacraments, creeds and sacred books, are viewed as paltry substitutes, which are doled out to those who, by reason of incapacity or lack of effort, miss the firsthand mystic communion; or else they are external aids that are of use only in the earlier stages of the quest, crutches to be thrown away when direct access to God is attained.

**RESPONSE TO THE DIVINE.** In sacramental religion, where the divine is apprehended chiefly in material embodiments, the center of religious activity will be found in ritual acts centering on these embodiments. The sacred places, animals, statues, and such, must be treated with reverence, approached and made use of with due precautions; and around these usages tend to grow prescribed rites. Since the sense of the divine presence in certain objects is likely to be enhanced by participation in solemn ceremonials centering on these objects, the religious activity becomes a self-perpetuating system, embodying what is currently called positive feedback.

In sacramental religion, the ritual tends to absorb most of the religious energies of the adherents and to

crowd the other elements out of the center of the picture. Primitive religion, which is strongly sacramental in character, is often unconcerned with moral distinctions; and we might speculate that the progressive moralization of religion is achieved at the expense of ritual preoccupations. We can see this conflict at many points in the history of religions, most notably in the denunciations that the Hebrew prophets directed against the ritual-minded religionists of their day, and in their exhortations to substitute thirst for righteousness for the concern for niceties of ceremony. Even in its highest developments, sacramental religion tends to slacken the ethical tension that is found in prophetic religion. Where sacramentalism is strong in a monotheistic religion, the natural tendency is to take everything in nature as a divine manifestation. If everything is sacred, then nothing can be fundamentally evil; and thus the distinction between good and evil becomes blurred. One of the elements in the Protestant Reformation was a protest against tendencies to blurring of this sort, which took place in the largely sacramental medieval form of Christianity.

The typical response of prophetic religion to the divine is also nicely coordinated with the chief form in which the divine is apprehended. The reaction naturally called for by a message from the divine is acceptance. This involves both an intellectual acceptance of its contents—belief that whatever statements it makes are true—and obedience to the commands and exhortations it contains. Hence, in prophetic religion faith is the supreme virtue, and affirmations and confessions of faith play an important role. This is illustrated by the insistence of such great Christian prophetic figures as Paul and Martin Luther on faith in Christ as both necessary and sufficient for salvation and by the Muslim practice of repeating daily the creed “There is no God but Allah, and Muḥammad is his prophet.” It is important to realize that faith in this sense means far more than the intellectual assent to certain propositions. It also involves taking up an attitude on the basis of that affirmation and expressing that attitude in action. The Jewish prophet Micah expressed the essence of prophetic religion when he said, “What doth the Lord require of thee, but to do justly and to love mercy, and to walk humbly with thy God?” Thus, it would not be incorrect to say that the emphasis in the prophetic response is ethical, providing we do not separate ethics from the believing acceptance of the divine message that is its foundation.

To understand the typical response of mystical religion, we must remember that for the mystic, immediate identification with the divine is of supreme importance.

Therefore he concentrates on an ascetic and contemplative discipline that will be conducive to the attainment and maintenance of that condition. He tends to become involved in abstentions and self-tortures designed to wean him from his attachment to things of this world, and in contemplative exercises designed to withdraw the attention from finite things, leaving the soul empty and receptive to influences from the divine. He will make use of ceremonies and will accede to moral principles insofar as he believes them to be efficacious in furthering his ultimate goal. But ultimately they must go; when union with God has been achieved, they are of no more significance. Thus, like sacramentalism, mysticism tends toward the amoral. Only rarely does either become completely amoral, and then for different reasons. For the sacramentalist, conventional moral distinctions may come to seem unimportant because he views everything as equally saturated with the divine; they seem unimportant to the mystic because every finite object or activity is outside the mystic union, and so all are, in the end, equally worthless. The righteous and the wicked are equally far from the true religious goal. While united with God, one does not act.

**PLACE OF DOCTRINE.** Finally, we may compare the three types of religion with respect to the status of beliefs and creeds. Since faith is central for prophetic religion and since the word is stressed as the primary medium of divine manifestation, it is not surprising that in prophetic religion, creed and doctrine are emphasized more than in the others. Mystical religion, at its purest, is indifferent to matters of belief and doctrine. The mystical experience and the divinity it reveals are often regarded as ineffable, not to be expressed in human language; hence, mystics tend to reject all doctrinal formulations as inadequate. At best, a mystic will admit that some formulations are less inadequate symbols of the unutterable than are others. Thus, in such predominantly mystical groups as the Sufis and the Quakers, little or no attempt is made to enforce doctrinal conformity. And in an extreme form of mysticism, like that of Zen Buddhism, any doctrinal formulation is discouraged. Sacramental religion occupies a middle ground in this respect. In its more primitive forms, it is often extremely indefinite about belief. It has been said that primitive man “dances out his religion.” Certainly the elaboration of ritual in primitive religion far outstrips the associated theory. The primitive will often possess an incredibly detailed set of ritual prescriptions but have only the haziest idea of what there is about the nature or doings of the gods that makes them appropriate. In its more developed forms, sacramental theology

becomes more definite, but it is still true that to the extent that a religion is preoccupied with a sacramental approach to the divine, it is more impatient than prophetic religion with doctrinal subtleties.

We can coordinate this classification with the list of religion-making characteristics by pointing out that sacramental religion stresses sacred objects and ritual, prophetic religion stresses belief and morality, and mystical religion places chief emphasis on immediate experience and feeling.

**CONCRETE APPLICATION.** When we come to apply our scheme to particular cases, we must not suppose that any religion will fall completely in one class or another. In fact, it is better not to think of types of religions, but of religious tendencies that enter in varying proportions into the makeup of any actual religion. However, we can usually say that one tendency or another predominates in a given religion. Thus, Buddhism and philosophical Hinduism are predominantly mystical; Judaism, Islam, and Confucianism are primarily prophetic; and popular Hinduism, in company with all polytheistic and primitive religions, is primarily sacramental. Often a religion that begins with a definite bent will admit other elements in the course of its development. Islam, which began as the most severely prophetic of religions, has developed one of the world's most extreme groups of mystics in the Sufis, who are completely out of harmony with the spirit of Muḥammad, no matter how they may continue to express themselves in his phrases. Again, in Tibet, Buddhism has undergone a development quite foreign to its founder's intentions, blossoming into an extremely elaborate sacramentalism.

Christianity furnishes a good opportunity to study the intermingling and conflict of the different tendencies. It began as an outgrowth of Jewish prophecy, but in the process of adapting itself to the rest of the Western world it took on a considerable protective coloration of both the sacramental and mystical, and these aspects have remained with it throughout its career. Christian mysticism presents a good example of an element existing in a religion that is dominated by another element. As the price of toleration, Christian mystics have had to pay lip service to the official theology and to the prophetic moral element; and as a result, mystic thought and practice in Christianity have seldom received the extreme development found in India. In those cases where the mystical spirit has burst the fetters, as with Meister Eckhart, official condemnation has often resulted.

Looking at Christianity today, it can be said that although it is predominantly a prophetic religion, as compared with Hinduism and Buddhism, with respect to its internal divisions the Catholic wing (both Roman and Greek) tends more toward the sacramental, while the Protestant is more purely prophetic, with mysticism appearing sporadically throughout. In Catholicism the elaborateness of prescribed ceremonies, the emphasis on the necessity of material sacraments for salvation, and the insistence on a special status for consecrated priests are all typically sacramental. In Protestantism the emphasis on the sermon (the speaking forth of the Word of God) rather than on ritual, the emphasis on the Bible as the repository of divine revelation, and the moral earnestness and social concern are all earmarks of the prophetic spirit. "Religion" new copy p. 235:

**See also** Buddhism; Chinese Philosophy: Religion; Christianity; Creation and Conservation, Religious Doctrine of; Epistemology, Religious; Islamic Philosophy; Jewish Philosophy; Philosophy of Religion, History of; Philosophy of Religion, Problems of; Philosophy of Religion; Religion and Morality; Religion and the Biological Sciences; Religion and the Physical Sciences; Religion, Naturalistic Reconstructions of; Religion, Psychological Explanations of; Religious Experience, Argument for the Existence of God; Religious Experience; Religious Language; Theism, Arguments For and Against.

### **Bibliography**

Friedrich Schleiermacher, *On Religion*, translated by John Oman (New York: Harper, 1958), and *The Christian Faith*, translated by H. R. Mackintosh (Edinburgh, 1956), contain classic statements of the view that religion is essentially a mode of experience. A more recent statement of this point of view that emphasizes mystical experience is in W. T. Stace, *Time and Eternity* (Princeton, NJ: Princeton University Press, 1952). The moral aspect of religion is stressed in Immanuel Kant, *Religion within the Limits of Reason Alone*, translated by T. M. Greene and H. H. Hudson (La Salle, IL: Open Court, 1934), and in John Baillie, *The Interpretation of Religion* (New York: Scribners, 1928). Conceptions of religion from the standpoint of philosophical naturalism are to be found in Auguste Comte, *A General View of Positivism* (London, 1865); Ludwig Feuerbach, *The Essence of Christianity*, translated by George Eliot (New York: Ungar, 1957); George Santayana, *Reason in Religion* (New York: Scribners, 1905); John Dewey, *A Common Faith* (New Haven, CT: Yale University Press, 1934); and Erich Fromm, *Psychoanalysis and Religion* (New Haven, CT: Yale University Press, 1950). Stimulating discussions, not so easily classified, are Henri Bergson, *The Two Sources of Morality and Religion*, translated by R. Ashley Audra (New York: Holt, 1935); William James, *The Varieties of Religious Experience* (New



York: Longman, 1902); Josiah Royce, *The Sources of Religious Insight* (New York: Scribners, 1912); John Oman, "The Sphere of Religion," in *Science, Religion and Reality*, edited by Joseph Needham (New York: Macmillan, 1925); and Alfred North Whitehead, *Religion in the Making* (New York: Macmillan, 1926). Important discussions from the social sciences include Émile Durkheim, *The Elementary Forms of the Religious Life*, translated by J. W. Swain (London: Allen and Unwin, 1915), and Bronisław Malinowski, *Magic, Science, and Religion* (Garden City, NY: Doubleday, 1954). James H. Leuba, *A Psychological Study of Religion* (New York: Macmillan, 1912), and Julian Huxley, *Religion without Revelation* (New York: Harper, 1958), provide extensive critical discussion of a wide variety of definitions, as well as presenting original conceptions.

#### OTHER RECOMMENDED TITLES

- Banton, Michael, ed. *Anthropological Approaches to the Study of Religion*. Reprint edition, London: Routledge, 1990.
- Batson, C. Daniel et al. *Religion and the Individual: A Social-Psychological Perspective*. New York: Oxford University Press, 1993.
- Berger, Peter. *The Social Reality of Religion*. Harmondsworth, U.K.: Penguin, 1967.
- Boyer, Pascal. *Religion Explained: The Evolutionary Origins of Religious Thought*, Reprint edition. New York: Basic, 2002.
- Braun, Willi, and Russell McCutcheon, eds. *Guide to the Study of Religion*. London: Cassell, 2000.
- Byrne, Peter. *Natural Religion and the Nature of Religion*. London: Routledge, 1989.
- Clarke, Peter, and Peter Byrne, eds. *Religion Defined and Explained*. New York: St. Martin's Press, 1993.
- Derrida, Jacques, and Gianni Vattimo, eds. *Religion*. Stanford, CA: Stanford University Press, 1998.
- Godlove, Terry. *Religion, Interpretation, and Diversity of Belief: The Framework Model from Kant to Durkheim to Davidson*. Cambridge, U.K.: Cambridge University Press, 1989.
- Guthrie, Steward. *Faces in the Clouds: A New Theory of Religion*. New York: Oxford University Press, 1993.
- Hick, John. *An Interpretation of Religion: Human Responses to the Transcendent*. New Haven, CT: Yale University Press, 1989.
- Lawson, E. Thomas, and Robert McCauley. *Rethinking Religion: Connecting Cognition and Culture*. New York: Cambridge University Press, 1990.
- Phillips, D. Z. *Religion without Explanation*. Oxford: Blackwell, 1976.
- Smart, Ninian. *The Science of Religion and the Sociology of Knowledge*. Princeton, NJ: Princeton University Press, 1973.
- Yinger, J. Milton. *The Scientific Study of Religion*. New York: Macmillan, 1970.

**William P. Alston (1967)**

*Bibliography updated by Christian B. Miller (2005)*

## RELIGION, NATURALISTIC RECONSTRUCTIONS OF

In philosophy a naturalist is one who holds that there is nothing over and above nature. A naturalist is committed to rejecting traditional religion, which is based on beliefs in the supernatural. This does not necessarily carry with it a rejection of religion as such, however. Many naturalists envisage a substitute for traditional religion that will perform the typical functions of religion without making any claims beyond the natural world. We can best classify naturalistic forms of religion in terms of what they take God to be—that is, what they set up as an object of worship. In traditional religion the supernatural personal deity is worshiped because he is thought of as the zenith of both goodness and power. More generally, we can say that religious worship is accorded to any being because it is regarded as having a controlling voice in the course of events and at least potentially exercising that power for the good. This suggests that to find a focus for religious responses in the natural world, we should look for a basic natural source of value. Forms of naturalistic religion differ as to where this is located. Broadly speaking, achievements of value in human life are due to factors of two sorts: (1) man's natural endowments, together with the deposit of his past achievements in the cultural heritage of a society, and (2) things and processes in nonhuman nature on which man depends for the possibility of his successes and, indeed, his very life. Most naturalists locate their religious object primarily on one or the other side of this distinction, although some try to maintain an even balance between the two.

The first factor is stressed most by those who are called religious humanists. This group includes Ludwig Feuerbach and Auguste Comte in the nineteenth century and John Dewey and Erich Fromm in the twentieth. Of these men Comte has been the most influential.

#### COMTE

In Comte's view, it is to humanity that the individual man owes everything that he is and has. It is because he shares in the general biological and psychological capacities of human nature that he is able to live a human life. And the men of a given generation are able to lead a fully human life because of the labors of their predecessors in building up their cultural heritage. Moreover, according to Comte, the service of humanity, in the many forms this can take, is the noblest ideal that could be proposed to an individual; and humanity, unlike an omnipotent God, needs this service. Thus, Comte proposed to set up a religion of

humanity with man, viewed as a unitary though spatiotemporally scattered being, as the object of worship.

Unlike many naturalists Comte was not at all vague about the detailed functioning of his proposed religion. He was impressed with the ritual structure of Roman Catholicism and took it as his model. For example, in the analogue of baptism, the sacrament of presentation, the parents would dedicate their child to the service of humanity in an impressive public ceremony. Public observances were to be reinforced by the regular practice of private prayer, on which Comte laid the greatest stress. A person was to pray four times daily, with each prayer divided into a commemorative and a purificatory part. In the first part one would invoke some great benefactor of humanity; by reflecting gratefully on his deeds, one would be inspired to follow his example, and one's love of humanity would thus be quickened. The purificatory part would give solemn expression to the noble desires thereby evoked; in it the individual would dedicate himself to the service of humanity. Other rituals included a system of religious festivals and a calendar of the saints of humanity that provided the material for the prayers on each day of the year.

Some idea of the religious fervor generated in Comte by the contemplation of humanity may be gained from this quotation from *A General View of Positivism*:

The Being upon whom all our thoughts are concentrated is one whose existence is undoubted. We recognize that existence not in the Present only, but in the Past, and even in the Future: and we find it always subject to one fundamental Law, by which we are enabled to conceive of it as a whole. Placing our highest happiness in universal Love, we live, as far as it is possible, for others: and this in public life as well as in private; for the two are closely linked together in our religion; a religion clothed in all the beauty of Art, and yet never inconsistent with Science. After having thus exercised our powers to the full, and having given a charm and sacredness to our temporary life, we shall at last be forever incorporated into the Supreme Being, of whose life all noble natures are necessarily partakers. It is only through the worship of Humanity that we can feel the inward reality and inexpressible sweetness of this incorporation. (p. 444)

Comte had considerable influence in his lifetime, and a few functioning parishes of his religion of humanity sprang up. They have not survived, however, and a revival in our time hardly seems likely. In the twentieth

century, reeling under the impact of two world wars and the hourly expectation of the death knell of civilization, we are not inclined to grow misty-eyed over humanity. Recent humanists have tended to be more critical in their reverence. The latest trend is to single out the more ideal aspects of man—his aspirations for truth, beauty, and goodness—for religious worship. Or the emphasis shifts from man as he actually exists to the ideals that man pursues in his better moments. Thus, in his book *A Common Faith*, John Dewey defines God as “the unity of all ideal ends arousing us to desire and action” (p. 42).

## DEWEY

Unlike Comte, Dewey has no interest in developing an organized naturalistic religion. It would seem that religious organization and religious ritual are too closely associated in his mind with the supernaturalism that he rejects. For Dewey the important thing is the religious quality that experience can assume under certain conditions. Any unification of the whole self around the pursuit of an ideal end is religious in quality. Dewey is emphatic in insisting that this is a quality, rather than a kind, of experience. Whenever a person is thoroughly committed to the pursuit of any ideal, be it scientific, social, artistic, or whatever, his experience attains the kind of fulfillment that has always been characteristic of what is most valuable in religion. According to Dewey, in traditional religion this quality has been encumbered and obscured by irrelevant trappings, particularly the theological dogma in terms of which it has been pursued. In the past, self-integration in the pursuit of the ideal has been thought of as service of God, unity with God, or submission to God's will. It is Dewey's conviction that the religious quality can be more effectively sought if the quest is not carried on under this banner. To reflective men, supernaturalistic dogma will always appear dubious at best. If the quest for self-integration in the service of the ideal is too closely tied to theology, it will be endangered when the theology is rejected as rationally groundless. Moreover, insofar as the theology is taken seriously, it diverts attention from the active pursuit of the ideal. Worse, the assurance that the good is already perfectly realized in the divine nature has the tendency to cut the nerve of moral effort; in that case it is not up to us to introduce the good into the world. Thus, Dewey's main concern as a philosopher of religion is to redirect religious ardor into the quest for a richer quality of human life rather than to construct a framework for a naturalistically oriented religious organization.

There is no developed naturalistic philosophy of religion that stresses the nonhuman side of the natural sources of value to the extent to which Comte stresses the human side. (Though we can find this in literature, notably in Richard Jeffries, who had a kind of religious intoxication with inanimate nature without, however, conceiving of it as suffused with a spiritual being or beings. This is a naturalistic counterpart of the nature worship of ancient Greece, just as Comte's religion of humanity is a naturalistic counterpart of an ethical monotheism like Christianity.) However, there is a marked tendency among contemporary naturalists to emphasize the nonhuman side much more than Comte or Dewey. Good examples of this are the liberal theologian Henry Nelson Wieman and the biologist Julian Huxley, who in his book *Religion without Revelation* has made the most coherent and comprehensive recent attempt to sketch out a naturalistically oriented religion.

#### HUXLEY

According to Huxley's conception, religion stems from two basic sources. One is man's concern with his destiny—his position and role in the universe and their implications for his activity; the other is the sense of sacredness. Following Rudolf Otto, Huxley thinks of the sense of sacredness as a unique kind of experience that is an intimate blend of awe, wonder, and fascination; this mode of feeling arises spontaneously in reaction to a wide variety of objects and situations. Religion, then, is a social organ for dealing with problems of human destiny. As such it involves a conception of the world within which this destiny exists, some mobilization of the emotional forces in man vis-à-vis the world thus conceived, some sort of ritual for expressing and maintaining the feelings and attitudes developed with respect to the forces affecting human destiny, and some dispositions with respect to the practical problems connected with our destiny. The sense of sacredness enters into the second and third of these aspects. As Huxley sees it, a way of dealing with problems of human destiny would not be distinctively religious if it did not stem from and encourage a sense of the sacredness of the major elements in its view of the world, man, and human life.

Huxley, as a thoroughgoing naturalist, holds that the supernaturalistic worldview in terms of which religion has traditionally performed its functions is no longer tenable in the light of modern scientific knowledge. Moreover, he thinks that it is possible to develop a full-blown religion on a naturalistic basis. As the intellectual basis for such a religion, Huxley puts forward "evolutionary natu-

ralism," a view of the spatiotemporal universe, inspired by modern biology and cosmology, in which the universe is conceived of as an indefinitely extended creative process, always tending to higher levels of development, with all the sources and principles of this creativity immanent in the process. The basic role of man is to be the chief agent of this evolutionary advance on earth through the application of his intelligence to the problems of life on Earth and through the building of a harmonious and stable community. A religion based on these conceptions will be focused on an object of worship that is a construct out of all the forces affecting human destiny, including basic physical forces as well as the fundamental facts of human existence and social life. God, then, will consist of all these factors, held together by the feeling of sacredness with which they are apprehended. As a start toward conceiving this assemblage as a unified object of worship, Huxley presents a naturalistic version of the Christian doctrine of the Trinity. God the Father is made up of the forces of nonhuman nature. God the Holy Ghost symbolizes the ideals toward which human beings at their best are striving. God the Son personifies human nature as it actually exists, bridging the gulf between the other two by channeling natural forces into the pursuit of ideals. And the unity of all the persons as one God represents the fact that all these aspects of the divine are intimately connected.

Many thinkers, atheists as well as theists, take a dim view of all these proceedings. Since the theists' lack of enthusiasm stems from obvious sources, let us concentrate on the atheists. The issues here are normative or evaluative rather than factual. Comte and Huxley as philosophers of religion are not, with perhaps minor exceptions, making any factual judgments with which other naturalists might disagree because they are making no factual judgments at all beyond their basic commitment to naturalism. If a man like Bertrand Russell or Jean-Paul Sartre disagrees with Huxley, he differs about the value of what Huxley is proposing. His low evaluation may have different bases. First, he may feel that man or the basic forces of nature constitute too pallid a substitute for the God of theism to afford a secure footing for the distinctively religious reactions of reverence, adoration, and worship. A man like Huxley might, for his part, interpret this as a reflection of a suppressed hankering after the old supernatural deity. Second, Russell or Sartre may turn this charge on Huxley and maintain that one searches for an object of worship within nature only because he has not sufficiently emancipated himself from the old religious orientation and that this religion of evolutionary naturalism represents an uneasy compromise between religious and secular orientations. It seems clear

that there is no one objective resolution of such disputes. People differ in such a way that different total orientations will seem congenial to people with different temperaments and cultural backgrounds. It is perhaps unfortunate, on the whole, that many people need to find something fundamentally unworthy in every other religion in order to find a firm attachment to their own religious positions, although it is undoubtedly true that religious discussions are more lively than they would be if this were not the case.

**See also** Comte, Auguste; Dewey, John; Evolutionary Theory; Feuerbach, Ludwig; Human Nature; Naturalism; Otto, Rudolf; Russell, Bertrand Arthur William; Sartre, Jean-Paul.

### **Bibliography**

- Ames, E. S. *Religion*. New York, 1929.
- Comte, Auguste. *A General View of Positivism*. Translated by J. H. Bridges. London, 1880. Ch. 6.
- Dewey, John. *A Common Faith*. New Haven, CT: Yale University Press, 1934.
- Feuerbach, Ludwig. *The Essence of Christianity*. Translated by George Eliot. New York: Harper, 1957.
- Fromm, Erich. *Psychoanalysis and Religion*. London: Gollancz, 1951.
- Huxley, Julian. *Religion without Revelation*. New York: Harper, 1957.
- Russell, Bertrand. *Why I Am Not a Christian and Other Essays*. New York: Simon and Schuster, 1957.
- Wieman, Henry Nelson. *The Source of Human Good*. Chicago: University of Chicago Press, 1946.

*William P. Alston (1967)*

## RELIGION, NATURALISTIC RECONSTRUCTIONS OF [ADDENDUM]

### CONTEMPORARY NATURALISTIC RELIGION

What should one contrast nature with? The supernatural, maybe? What is meant here by supernaturalism is the thesis that the divine is different in kind from familiar things and persons; and/or that there are divine interventions that are contrary to the laws of nature. If this is the relevant contrast then naturalistic religion requires merely that God be taken as either a person or a community of persons. God is then like humans, although infinitely more powerful, and acts in the world in whatever way people act when they exercise their freedom. Such anti-

supernaturalism is weaker than naturalism as understood by contemporary philosophers, who would balk at calling the dualist Richard Swinburne (2004), the idealists Timothy Sprigge (1983) and John Foster (2004), or even the nonreductive physicalist Peter Forrest (1996) naturalists. This suggests that naturalism is to be contrasted not merely with the supernatural but also with anthropocentric *Metaphysics*, which takes consciousness and agency as fundamental features of reality that may be used to explain but must themselves be accepted without explanation. Naturalism in this strong sense is unlikely to support the humanist attitudes of Auguste Comte or John Dewey, but coheres well with Julian Huxley's evolutionary naturalism.

### DEEP ECOLOGY

The most widespread contemporary naturalistic movement with religious tendencies is deep ecology, which typically goes beyond an attitude of aesthetic appreciation of—and scientific interest in—life on earth, to attitudes of reverence and self-sacrifice (Naess 1989). Combined with a suitable metaphysical system this could be a genuinely naturalistic religion, although neopagan movements such as Wicca tend to incorporate belief in the supernatural. Two such metaphysical systems are process theology and pantheism.

### PROCESS PHILOSOPHY

The process philosophy of Alfred North Whitehead, Charles Hartshorne, and most recently David Ray Griffin (2001) can support either a liberal theistic religion or deep ecology. Process philosophy counts as naturalistic because it is biocentric rather than anthropocentric, in that it relies on preconscious sensitivity to the environment (prehension) and final causation. For that reason the God of process philosophy is immanent in the processes of the natural world, resulting in something similar to, although less austere than, Huxley's evolutionary naturalism. A chief objection to process philosophy is that we no longer have a theoretical need for either prehension or final causes even in biology.

### PANTHEISM

The universe as a whole or, perhaps better, the natural order is sufficiently awe-inspiring to ground some religious attitudes. So pantheism can form the basis of a naturalistic religion (Levine 1994). Like any religion this has metaphysical commitments: either the existence of the universe as a whole or the existence of laws of nature, but neither of these commitments would worry most naturalists.

## THE AFTERLIFE

Much religious motivation (for good and ill) lies in the belief in an afterlife. Does naturalism cohere with this belief? Granted that if there is a God concerned about individuals then there is not much problem, for there are ways God could ensure an afterlife without miracles and without there being souls (van Inwagen 1992). However, a pantheist God that just is the natural order will not be concerned about individuals, whereas the God of process philosophy might well lack the power required to be providential. Frank J. Tipler (1996) has suggested that in the distant future sentient beings will be able to reconstitute all the lives of those who have died. In his version all possible lives seem to get reconstituted, which prevents any of them being the same as early twenty-first century people. But one might surmise that there are traces of actual lives that could be used to reconstitute only those who have actually lived. A less far-fetched naturalistic account of the afterlife is based on the many worlds interpretation of quantum theory. For if there are many parallel universes and every physically possible event occurs in some of them, then in some of them it seems humans survive anything (Price 1996, ch. 9; Lewis 2004) The chief problem with such scenarios is over-survival, that is, at each moment each person divides into millions of successors.

## CONCLUSION

Not surprisingly the more narrowly naturalism is understood the more drastic a naturalistic reconstruction of religion must be. At one extreme, anti-supernaturalism sits comfortably with all but conservative religious movements. At the other, naturalists might reject even the biocentrism of process thought and be left with only a rather austere pantheism.

**See also** God, Concepts of; Naturalism; Pantheism; Physicalism; Whitehead, Alfred North.

## Bibliography

- Forrest, Peter. *God without the Supernatural*. Ithaca, NY: Cornell University Press, 1996.
- Foster, John. *The Divine Lawmaker: Lectures on Induction, Laws of Nature, and the Existence of God*. Oxford: Clarendon Press, 2004.
- Griffin, David Ray. *Reenchantment without Supernaturalism*. Ithaca, NY: Cornell University Press, 2001.
- Levine, Michael P. *Pantheism: A Non-Theistic Concept of Deity*. New York: Routledge, 1994.
- Lewis, David. "How Many Lives Has Schrödinger's Cat?" *Australasian Journal of Philosophy* 82 (2004): 3–22.

- Naess, Arne. *Ecology, Community, and Lifestyle: Outline of an Ecosophy*. Translated and revised by David Rothenberg. Cambridge, U.K.: Cambridge University Press, 1989.
- Price, Huw. *Time's Arrow & Archimedes Point: New Directions for the Physics of Time*. Oxford: Oxford University Press, 1996.
- Sprigge, Timothy. *The Vindication of Absolute Idealism*. Edinburgh, Scotland: Edinburgh University Press, 1983.
- Swinburne, Richard. *The Existence of God*. 2nd ed. Oxford: Clarendon Press, 2004.
- Tipler, Frank J. *The Physics of Immortality: Modern Cosmology, God and the Resurrection of the Dead*. London: Pan, 1996.
- van Inwagen, Peter. "The Possibility of Resurrection." In *Immortality*, edited by Paul Edwards, 242–246. New York: Macmillan, 1992. Reprinted from the *International Journal for the Philosophy of Religion* 9 (1978).

Peter Forrest (2005)

## RELIGION, PSYCHOLOGICAL EXPLANATIONS OF

In the seventeenth and eighteenth centuries the chief impact of science on religion came from the revised picture of the cosmos that emerged from developments in astronomy and physics. In the nineteenth century the impact was from the changed view of the history of life on Earth that was presented by geology and evolutionary biology. In the twentieth century the social sciences had the greatest impact on religion, although of a different nature. Physics and biology worried theologians because they introduced theories about the cosmos, life, and man that were at variance with beliefs intimately bound up with the religious tradition, such as the special creation of man. The impact of the social sciences, on the other hand, comes not from theories that contradict basic religious doctrines but from explanations of religion itself that seem to rob it of its significance.

Since the nineteenth century numerous ideas have been put forward as to the psychological and sociological factors that are responsible for religion. The most important of these are (1) the Marxian theory that religion is one of the ideological reflections of the current state of economic interrelations in a society; (2) the similar, but more elaborately developed, theory of the sociologist Émile Durkheim that religious belief constitutes a projection of the structure of society; and (3) the Freudian theory that religious belief arises from projections designed to alleviate certain kinds of unconscious conflict. These are all scientific explanations in that they trace religion to factors wholly within the world of nature, and hence they

are, at least in principle, subject to empirical test. Concentration on one of these, the Freudian, will enable us to illustrate the philosophical problems raised by such explanations.

### THE FREUDIAN EXPLANATION

The Freudian account begins with certain similarities between attributes of and attitudes toward a personal deity, on the one hand, and the small child's conception of and mode of relating to his father, on the other. In both cases the superior being is regarded as omnipotent, omniscient, inscrutable, and providential. In both cases the individual reacts to this superior being with utter dependence, awe, fear of punishment, and gratitude for mercy and protection. These parallels suggest, though they do not prove, that the original model for the conception of God is to be found in the infantile conception of one's parents, and that the almost universal inclination to believe in personal deities is to be traced to psychological remnants of the infantile situation. According to Sigmund Freud, these remnants are mostly the result of the Oedipal conflict. According to his theory, around the age of four the boy (restricting ourselves to the male for simplicity of exposition) comes to desire his mother sexually and to regard his father as a rival. Reacting more or less to actual indications, the boy becomes so afraid of the father's hostility, and also so afraid of losing his love, that he not only abandons his sexual aims but also represses the entire complex of desires, fears, and conceptions. This complex remains, in greater or lesser intensity, in the unconscious; and it is because a supernatural personal deity provides an external object on which to project it that men have as much inclination as they do to believe in such a being and to accept the attitudes and practices that go with this belief.

To understand what the projection does for the individual, we must recognize that the repressed material involves severe conflict between tendencies to rebel against the father and tendencies to submit to the father, and between the Oedipal desires and the standards that would be violated by satisfying those desires. Projection of this material onto an external deity reduces distress in several ways. First, the externalization of the problem provides some relief. Instead of being plagued by mysterious discomfort, the individual is faced with a clear-cut opposition between various desires of his own and a forbidding external person. Second, there is less conflict because the external figure is so powerful as to seriously weaken the rebellion, and he is so idealized as to render resentment and hostility less appropriate. Third, there are

various mechanisms provided for dissipating the guilt over sexual desire for the mother and hostility toward the father. Confession, penance, and renunciations of various kinds afford socially approved means for relieving this guilt and counteracting its crippling influence.

People are more receptive to religious belief at some times than at others. Freud explains this in terms of the mechanism of regression. When a person encounters severe difficulties and frustrations at one stage of life, he tends to regress psychologically to an earlier stage at which these problems did not exist. Thus, when an adult is particularly hard pressed, there is generally some reinstatement of earlier modes of thinking, feeling, and relating to the environment. This means that the Oedipal material in the unconscious will become more intense and closer to the surface, while at the same time the person is more likely to engage in the childish practice of projection.

Thus, according to Freudian theory, an individual's tendency to accept belief in a supernatural personal deity (together with the other aspects of religious activity and involvement) is at least partly caused by a tendency to project a childhood father image existing in the unconscious, this projection normally following a regression set off by a current problem of adjustment and serving to alleviate unconscious conflicts and unconscious guilt. It is clear that, at best, this is only a partial explanation of religious belief. For one thing, it presupposes the prior existence of the religious ideas in the culture; at most, it is an explanation of the individual's readiness to accept these ideas when they are proffered.

Freud tried to supply this lack by developing a parallel theory of the development of religion in society. According to this theory, religion develops as a projection of a psychological complex that results from unconscious racial memories of a primal murder of the tyrannical father figure of a "primal horde." Cultural development is thus treated along the same lines as the development of the individual; something like a "collective unconscious" is posited in which psychic material can be transmitted in an unconscious form from one generation to another. However, these ideas have never won any considerable degree of acceptance, and in discussing Freud we can concentrate on his account of the psychological basis of religion in the individual.

### CRITICISM OF FREUDIAN EXPLANATION

With respect to any scientific explanation of religion, there are two questions to be raised. (1) What reason is

there to accept it? (2) If it is true, what bearing does it have on the truth, value, or justifiability of religion? It is the second question that specially lies within the province of the philosophy of religion.

It is clear that the Freudian explanation does not imply that the beliefs of religion are false; Freud himself recognized this, though not all Freudians do. But it is often assumed that the success of any explanation of religion in terms of factors within the natural world would show that we do not need to bring anything supernatural into the explanation, and hence would seriously weaken religion's claims to credibility. However, this depends on how these claims were made. If religion is based solely on divine revelation, then the fact that we can give an adequate explanation of religion without bringing in divine activity, revelatory or otherwise, seriously affects—though it does not conclusively disprove—the claim that certain beliefs are true because they are communicated to man by God. But if rational arguments are advanced in support of religious doctrine, such as the classical arguments for the existence of God, then whatever force these arguments have is in no degree lessened by the fact—if it be a fact—that the psychological basis for religion is as Freud supposed. Of course, if the Freudian mechanisms constitute a necessary as well as sufficient condition of religious belief, then it follows that no one has any good reason for these beliefs. If anyone did have a good reason, that would itself be a sufficient condition of the belief, and this would show that it is possible to have the belief without needing to project an unconscious father image. However, it is almost inconceivable that we should show that projection is a necessary condition of belief. At most, we could hope to show that there is some correlation between degree of unconscious Oedipal conflict and firmness of religious belief. Showing that a certain set of natural factors is one of the things that can produce religious belief may well nullify certain ways of supporting the beliefs, but it could hardly show that no adequate rational grounds could be produced.

There is another way in which it has been thought that the Freudian theory of religion carries with it a negative evaluation of religion. The particular causal factors to which Freud traced religion are of a sort associated with undesirable patterns of organization. To regard religion as caused by these factors is to class it with neurotic and infantile modes of behavior, and as such it is hardly worthy of serious consideration. In this respect, too, the psychoanalytic explanation is typical. One can imagine an explanation that traces religious activity to evaluatively neutral natural factors, such as patterns of neural

activity in the brain, but all the explanations in the field trace religion to states and activities that are more or less irrational, immature, or unworthy. Projection is involved in all the theories cited at the beginning of this article; the Marxist theory adds the point that religion is used by the dominant class to provide illusory consolations to those being exploited.

To be clear on this issue, we must distinguish the different forms these claims can take. Psychoanalytic literature is often simply an enumeration of similarities between religion and compulsion neuroses, such as firm attachment to rituals without having a rational explanation of the attachment. However, the similarity in itself proves nothing. A scientist “obsessed with an idea” also exhibits marked similarities to a compulsion neurotic, but this has no implications for the value of his work. The more important claim has to do with the causal factors said to underlie religion. Here, too, we must distinguish between (1) the claim that some neurotic condition is always or generally among the factors producing attachment to a religion, and (2) the claim that the causal basis of such attachment is markedly similar to the basis of recognized neuroses. There is no real evidence for the first claim. Controlled studies on the required scale have never been carried out. As for the second, we must ask how similar the causal basis is and what implications we are to draw from whatever degree of similarity exists. The mere fact that religion involves projection as a relief from unconscious conflict is not sufficient ground for labeling religion, in Freud's terms, “the universal obsessional neurosis of mankind.” We must distinguish between pathological and healthy resolutions of unconscious conflict.

The anti-Freudian psychoanalyst Carl Jung, in terming religion an alternative to neurosis, expressed his belief that it is a healthy outcome. The basic issue involved here concerns the definition of “neurosis.” If we define it in terms of a certain causal basis, then it may be that according to the Freudian theory, religion is, by its very nature, a form of neurosis. But then it remains an open question whether or not it is a desirable, justifiable, or realistic mode of activity. If neurosis is defined in this way, we may have to distinguish between good and bad neuroses. If, on the other hand, we accept common usage and build a negative evaluation into the definition of neurosis (by having as a necessary condition of neurosis that it make a satisfactory adjustment to one's environment difficult), then it would no longer be an open question whether religion, if neurotic, is a good thing. But with this concept of neurosis, we have a much stronger thesis, which calls for evidence that has not yet been pro-

vided. No one has shown that in general religious believers are less able to establish satisfying personal relations and less able to get ahead in their work than are nonbelievers. Even if this were shown, there would be further problems of a very sticky sort. The believer might complain that restricting “the environment” to the natural environment is question-begging. He would say that whatever the bearing of religious attachment on getting along in human society, it is essential to adequate adjustment to God and his demands. To ignore this aspect of “the environment” is to employ a criterion of adjustment that presupposes the falsity of religious beliefs.

Similar comments apply to the idea that the psychoanalytic theory implies that religion is infantile and hence unworthy of mature men. It is true that the way a religious man relates himself to God is in many ways similar to the way a small child relates himself to a father. But whether or not this is a mature, realistic mode of activity is wholly a function of whether there really is such a God. If there is, then this is the only reasonable stance to take. Hence, to condemn religion on these grounds is to presuppose the falsity of its beliefs.

Thus, there are many gaps in any line of reasoning that tries to derive a negative evaluation of religion from a causal explanation of religion in psychological or sociological terms. If a person does not feel that he has a firm basis for his religious beliefs, then looking at religion in a Freudian or Marxian light may well lead him to give up his beliefs. More generally, we can say that Freudian or Marxian theory does not provide an intellectual atmosphere in which one would expect religious belief to flourish; but it does not appear that these theories, as so far developed, are in any way logically incompatible with the truth, justifiability, and value of traditional religion.

**See also** Durkheim, Émile; Freud, Sigmund; Jung, Carl Gustav; Marxist Philosophy; Philosophy of Religion, Problems of; Popular Arguments for the Existence of God.

### **Bibliography**

Important treatments of religion from a psychological point of view include Sigmund Freud, *Totem and Taboo*, translated by A. A. Brill (New York: Moffat, Yard, 1918), *The Future of an Illusion*, translated by W. D. Robson-Scott (New York: Liveright, 1928), and *Moses and Monotheism*, translated by Katherine Jones (New York: Knopf, 1939); Carl Jung, *Psychology and Religion* (New Haven, CT: Yale University Press, 1938), and *Modern Man in Search of a Soul*, translated by W. S. Dell and C. F. Baynes (New York: Harcourt Brace, 1933); Erich Fromm, *Psychoanalysis and Religion* (New Haven, CT: Yale University Press, 1950); A. T. Bosien, *The*

*Exploration of the Inner World* (New York: Willett, Clark, 1952); and Theodor Reik, *Dogma and Compulsion*, translated by Bernard Miall (New York: International Universities Press, 1951). R. S. Lee, *Freud and Christianity* (New York: A. A. Wyn, 1949), and William P. Alston, “Psychoanalytic Theory and Theistic Belief,” in *Faith and the Philosophers*, edited by John Hick (New York: St. Martin’s Press, 1964), present discussions of the Freudian treatment of religion.

For a sociological point of view, see Émile Durkheim, *The Elementary Forms of the Religious Life*, translated by J. W. Swain (London: Allen and Unwin, 1915); Vilfredo Pareto, *The Mind and Society*, translated by Andrew Bongiorno and Arthur Livingston (New York: Harcourt Brace, 1935), Vol. III; V. F. Calverton, *The Passing of the Gods* (New York: Scribners, 1934); and G. E. Swanson, *The Birth of the Gods* (Ann Arbor: University of Michigan Press, 1960).

**William P. Alston (1967)**

## RELIGION, PSYCHOLOGICAL EXPLANATIONS OF [ADDENDUM]

During the last few decades of the twentieth century scientific ability to explore the brain directly increased dramatically, so neuroscientific discoveries during the period resulted in a broadening of perspectives from which psychological explanations of religion may be given. First, the ideological impasse on method between behavioristic and psychoanalytic or introspective approaches in psychology yielded to more pragmatic heterophenomenological (Dennett 2003) or neurophenomenological (Varela, Thompson, and Rosch 1991) methods for investigating mental states. Second, Platonic and Cartesian views of emotion as inherently irrational and subversive of productive cognitive functioning were contested by studies that showed that absence of emotion produced a cognitively dysfunctional Phineas Gage, not a pure-minded Philosopher King (Damasio 1994). Third, the Enlightenment notion of a person as an isolated, autonomous rational optimizer, a “ghost” in a bodily machine, began to yield to a notion of a person as an embodied and interactive global workspace (Baars 1997) that is distributed across both interpersonal relationships (attachment theory; Panksepp 1998) and the environment (Clark 1999). Fourth, clinical, cognitive, and transpersonal psychologists (Wilber 1998) began to see the value of studying and using religion in their clinical practices to aid in communication, understanding, and healing.



The net effect of these shifts in perspective on psychological methodology and ontology has been a return to a Jamesian (James 1981) view of human psychology as consisting of a stream of variously conscious and unconscious processes, related to one another in modular ways (Fodor 1983; Weiskrantz 1997), and integrated somewhat haphazardly through the accidents of evolutionary history.

New avenues of exploration for religious psychological states, beliefs, and practices have been opened by these developments in cognitive neuroscience, as well as by new technology. Some of these include:

- (1) Brain scans: studies of the brains of persons engaged in religious activities, through Positron Emission Tomography, Computed Axial Tomography, and Single Photon Emission Computed Tomography (SPECT) scans, and a comparison of the experimental results with base-line brain scans and scans of persons with known pathological conditions such as brain lesions, schizophrenia, and epilepsy.
- (2) First-person methods: Without behaviorist presuppositions, methods for systematic and controlled introspection can be studied in a critical but open-minded way.
- (3) Health and integration studies: Studies of the interpersonal and integrative effects of religious experiences, beliefs, or practices are being done in clinical settings.

While it is still possible that the pathologies attributed to religious experience, beliefs, and practices by eliminativists, Freudians, Marxists, and Durkheimians might be corroborated by the emerging twenty-first-century evidence, religious psychology is now at least open to the vindication of religion from charges of pathology. In what follows, samples of each of the previous lines of inquiry into the psychology of religion in cognitive science are cited.

### BRAIN SCAN STUDIES

Brain scans of advanced-level meditators, persons suffering from hallucinations, and persons engaged in prayer or other religious ceremonies are being produced by researchers at a variety of universities and institutes. Michael A. Persinger (1993) induced hallucinations in laboratory subjects through stimulation of temporal lobes of the brain. Based on this evidence and reports of religious experience by schizophrenic and epileptic patients, he argues that religious experiences, as hallucina-

tions, are a result of kindling, erratic neuronal stimulation that spreads through sections of the brain. He also reports that enhanced geomagnetic activity and limbic seizures produce religious senses of a “felt presence” and that meditation contributes to intrusive experiences.

In contrast, Eugene G. d’Aquili and Andrew B. Newberg (1999) offer SPECT scans of advanced-level meditators that show changes in regional cerebral blood flow as evidence that alternate circuits of brain activity are developed during meditation. D’Aquili and Newberg discovered that during meditation there is increased activity in the frontal lobes of the brain correlated with decreased activity in the posterior parietal lobes of the brain. They claim that the result is deafferentation of the outward orientation and association areas of the prefrontal cortex, resulting in senses of spacelessness, timelessness, and selflessness typically associated with religious experiences that they characterize as Absolute Unitary Being experiences.

### FIRST-PERSON AND INTROSPECTIVE STUDIES

Neurophenomenologists are examining systematic approaches to introspection as a tool of study, using both Husserlian phenomenological techniques and meditative techniques developed in Asian religious traditions, to gain insight into the psychology of religious states of consciousness. The Mind and Life Institute, working at the Keck Laboratory at the University of Wisconsin, Madison, and at the University of Paris, engages in collaborative research between Buddhist meditators and Western neuroscientists, aimed at correlating the Buddhist first-person trained experience of focused attention, open attention, visualization, and compassion, with states of neural phase-symmetry detected on high-density electroencephalography, magnetoencephalography, and functional magnetic resonance imaging. The researchers hope to show that stabilized, trained, first-person experiences of focused attention, compassion, and so on can be systematically correlated to states of neural phase-synchrony that represent states of large-scale integration within the brain.

### HEALTH-INTEGRATION STUDIES

Psychologists, such as Mihaly Csikszentmihaly (1997), are studying the relationship between happiness and peak experiences of the type outlined by Abraham Maslow, and are discovering that highly engaged attitudes and relationships, of the type long encouraged by religions, are productive of happiness, or Aristotelian eudaimonia.

In this research self-sacrificing and loving relationships to work and significant others are turning out to produce both happiness and physical and mental health, despite predictions to the contrary made by psychological survivalist and egocentrist theories.

Research groups such as the John Templeton Foundation, the Metanexus Institute, and Stephen G. Post's Institute for Research on Unlimited Love are using methodologies that could be characterized as heterophenomenological to explore the health and social effects of compassionate behavior on human thriving. Also, Division 36 of the American Psychological Association has been sponsoring conferences, several journals (i.e., *International Journal for the Psychology of Religion*, *Journal for the Scientific Study of Religion*, and *Review of Religious Research*), and a newsletter cataloging its members' study of a wide variety of issues related to the clinical and psychological roles of religion in the family, in coping with illness and death, in youth violence, in gender studies, in psychotherapy, in shaping values, and in sociological group formation, among many other topics.

**See also** Mysticism, Nature and Assessment of; Religious Experience; Religious Experience, Arguments for the Existence of God.

### Bibliography

- Alper, Matthew. *The "God" Part of the Brain: A Scientific Interpretation of Human Spirituality and God*, 4th ed. Brooklyn, NY: Rogue Press, 2001.
- Andresen, Jensine, and Robert K. C. Forman, eds. *Cognitive Models and Spiritual Maps: Interdisciplinary Explorations of Religious Experience*. Thorverton, U.K.: Imprint Academic, 2002.
- Baars, Bernard J. *In the Theater of Consciousness: The Workspace of the Mind*. New York: Oxford University Press, 1997.
- Baars, Bernard J., William P. Banks, and James B. Newman, eds. *Essential Sources in the Scientific Study of Consciousness*. Cambridge, MA: MIT Press, 2003.
- Clark, Andy. *Being There*. Cambridge, MA: MIT Press, 1997.
- Csikszentmihaly, Mihaly. *Finding Flow: The Psychology of Engagement with Everyday Life*. New York: Basic Books, 1997.
- Damasio, Antonio R. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam, 1994.
- D'Aquili, Eugene G., and Andrew B. Newberg. *The Mystical Mind: Probing the Biology of Religious Experience*. Minneapolis, MN: Fortress Press, 1999.
- Dennett, Daniel C. "Who's on First? Heterophenomenology Explained." *Journal of Consciousness Studies* 10 (9–10) (2003): 19–30.
- Fodor, Jerry A. *The Modularity of Mind*. Cambridge, MA: MIT Press, 1983.
- Forman, Robert K. C., ed. *The Innate Capacity, Mysticism, Psychology, and Philosophy*. New York: Oxford University Press, 1998.
- Freedman, Anthony, ed. *The Emergence of Consciousness*. Thorverton, U.K.: Imprint Academic, 2001.
- Hood, Ralph W., Jr., et al. *The Psychology of Religion: An Empirical Approach*, 2nd ed. New York: Guilford Press, 1996.
- James, William. *The Principles of Psychology*. Edited by Frederick Burkhardt. Cambridge, MA: Harvard University Press, 1981.
- Lutz, Antoine, and Evan Thompson. "Neurophenomenology: Integrating Subjective Experience and Brain Dynamics in the Neuroscience of Consciousness." *Journal of Consciousness Studies* 10 (9–10) (2003): 31–52.
- Metanexus Institute, <http://www.metanexus.net>, 2004.
- Mind and Life Institute. "Training and Studying the Mind: Toward and Integration of Buddhist Contemplative Practices and Neurosciences." <http://www.mindandlife.org/collaboration.html>, 2002.
- Newberg, Andrew, et al. "The Measurement of Regional Cerebral Blood Flow during the Complex Cognitive Task of Meditation: A Preliminary SPECT Study." *Psychiatric Research: Neuroimaging Section* 106 (2) (2001): 113–122.
- Nielsen, Michael E. "Psychology of Religion in the USA." *Psychology of Religion*, <http://www.psywww.com/psyrelig/USA.html>, February 2000.
- Núñez, Rafael, and Walter J. Freeman. *Reclaiming Cognition: The Primacy of Action, Intention, and Emotion*. Thorverton, U.K.: Imprint Academic, 1999.
- Panksepp, Jaak. *Affective Neuroscience: The Foundations of Human and Animal Emotions*. New York: Oxford University Press, 1998.
- Pargament, Kenneth I. *The Psychology of Religion and Coping: Theory, Research, Practice*. New York: Guilford Press, 1997.
- Persinger, Michael A. "Transcendental Meditation and General Meditation Are Associated with Enhanced Complex Partial Epileptic-Like Signs: Evidence for Cognitive Kindling?" *Perceptual and Motor Skills* 76 (1) (February 1993): 80–82.
- Varela, Francisco J., and Jonathan Shear. *The View from Within: First-Person Approaches to the Study of Consciousness*. Thorverton, U.K.: Imprint Academic, 2000.
- Varela, Francisco J., Evan Thompson, and Eleanor Rosch. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press, 1991.
- Weiskrantz, Lawrence. *Consciousness Lost and Found: A Neuropsychological Exploration*. New York: Oxford University Press, 1997.
- Wilber, Ken. *The Essential Ken Wilber: An Introductory Reader*. Boston: Shambhala Publications, 1998.

Laura E. Weed (2005)

## RELIGION AND MORALITY

Morality is closely associated with religion in the minds of many people. When religious leaders speak out on moral topics, their opinions are often treated with special

deference. They are regarded as moral experts. This raises the question of whether morality depends in some way on religion. Many philosophers have held that it does. John Locke, for example, argued that atheists could not be trusted to be moral because they would not consider themselves obliged even by solemn oaths, much less by ordinary promises. The answer to this question may be of considerable practical importance. If morality does depend on religion, the process of secularization, in the course of which religious belief and practice wither away, seems to pose a serious threat to morality. At one time many social theorists were confident that secularization was inevitable in modern and postmodern societies. Experience has undermined this confidence. Secularization no longer appears to be an inevitable consequence of modernization. Moreover, the process seems to occur at different rates in different modern societies. Thus secularization is more advanced in some Western European societies than it is in the United States. Nevertheless, it seems reasonable to be concerned about whether morality will decline to the extent that modern societies become more secular if it is the case that morality depends on religion.

This entry discusses several ways in which morality may depend on religion. It considers causal, conceptual, epistemological, and metaphysical dependency relations. It also explores the possibility that morality and religion may come into conflict. But a fruitful discussion of how two things are related must rely on some understanding of what those two things are. Hence the entry begins with characterizations of domains of morality and religion.

## MORALITY AND RELIGION CIRCUMSCRIBED

Understood in broad terms, morality consists of answers to the general normative question of how one should live one's life. It covers a wide range of topics related to the conduct of human life. Morality concerns actions that should and should not be performed and rules of conduct that should and should not be followed. It also comprehends motives for actions that people should and should not have and character traits or habits that people should and should not try to develop. Another subject of moral concern is ideals of saintliness or heroism to which some people may properly aspire, even though not everyone is called upon to live up to these ideals. Yet another subject is social and political arrangements that people should and should not strive to create or to sustain. Thus understood, morality consists of a diverse array beliefs and practices, and it is probably not possible to give an

illuminating definition of its scope. Philosophers often say that the realm of morality in this broad sense coincides with the realm of the ethical.

When philosophers reflect on the contents of the ethical, they find it useful to distinguish within it two domains, each characterized by a distinctive family of fundamental concepts. One is the axiological domain. Its basic concepts are goodness, badness, and indifference. The other is the deontological domain. Its basic concepts are requirement (obligation), permission (rightness), and prohibition (wrongness). Duty is the chief subject matter of the deontological domain. Some philosophers—Bernard Williams, for example—have proposed that morality be conceived narrowly as restricted to the deontological domain. On this conception, the domain of morality is a proper subdomain of the realm of the ethical.

Discussions of whether morality depends on religion frequently focus exclusively on the deontological domain. It is not hard to see why this occurs. Deontology consists of a system of requirements, permissions, and prohibitions. It is structurally similar to systems of law. Hence it is natural to think of deontology as the domain of moral law. Once this way of thinking has been adopted, the question arises as to whether moral law's binding force depends on the authority of a divine lawgiver. Most of the discussion in this entry will address the issue of whether moral requirements (obligations) and prohibitions (wrongness) depend on a deity of the sort to which the major monotheisms of Judaism, Christianity, and Islam are committed. However, some consideration will also be given to the topic of whether axiological goodness depends on such a deity. For this reason, the narrow conception of morality—which restricts it to the deontological domain—will not be adopted in this entry.

Religion, too, consists of beliefs and practices that exhibit great diversity. Most scholars who study it doubt that the concept of religion can be defined or analyzed in terms of necessary and sufficient conditions for being a religion. Some philosophers—for instance, John Hick—take the concept of religion to be a family-resemblance concept. On this view, religions resemble one another as members of a family resemble one another. For example, the ancient cults of Moloch, Christianity, and Theravada Buddhism may be classified as religions because they resemble one another in various respects, without supposing that all three of them satisfy a single set of necessary and sufficient conditions for being a religion. A more refined version of this view is provided by accounts developed in cognitive psychology of concepts organized

around examples that serve as prototypes or paradigms. As a result of complex patterns of similarity to—and difference from—the prototypes, other cases lie at various distances from the prototypes in a similar space. Cases near the prototypes fall under the concept; cases far enough away from the prototypes do not fall under the concept. In between there may be a gray area in which can be found borderline cases.

In attempting to define the concept of religion in terms of necessary and sufficient conditions, there is often disagreement about whether commitment—in theory or in practice—to superhuman beings is a necessary condition for being a religion. A celebrated debate in anthropology nicely illustrates such disagreement. Melford Spiro made the following proposal: “I shall define ‘religion’ as ‘an institution consisting of culturally patterned interaction with culturally postulated superhuman beings’” (Spiro 1966, p. 96). However, there is an obvious objection to Spiro’s proposal. In its purest form, Theravada Buddhism does not postulate superhuman beings. Yet most scholars think that pure Theravada Buddhism counts as a religion. So Spiro’s proposal fails to provide an adequate necessary condition for being a religion. It is too narrow.

Clifford Geertz (1966) offered a more complex definitional proposal. According to Geertz, “a *religion* is: (1) a system of symbols which acts to (2) establish powerful, pervasive, and long-lasting moods and motivations in men by (3) formulating conceptions of a general order of existence and (4) clothing these conceptions with such an aura of factuality that (5) the moods and motivations seem uniquely realistic” (Geertz 1966, p. 4). Theravada Buddhism will count as a religion by this definition. But so too will the system of symbols characteristic of Nazism, although most scholars wish to classify Nazism as a secular political ideology rather than as a religion—or at least to insist that it is religious only in some extended or analogical sense. Thus Geertz’s proposal fails to provide an adequate sufficient condition for being a religion. It is too broad. Disagreements of this kind fuel skepticism about whether it is possible to frame an illuminating definition of the concept of religion in terms of necessary and sufficient conditions.

For historical reasons, the monotheistic religions of Judaism, Christianity, and Islam are the prototypes of religion for people brought up within European and North American cultures. Discussion in this entry will focus almost entirely on the theism that is common to these paradigmatic religions.

## CAUSAL DEPENDENCE

Morality would depend historically on religion if moral beliefs and practices were derived by causal processes from prior religious beliefs and practices. It is often imagined that early human societies had worldviews in which no distinctions were drawn between moral and religious beliefs and practices. All norms of human conduct were then religious in character; their authority was taken to rest on superhuman sources such as the prescriptions of gods. Independent moral beliefs and practices emerged from such religious worldviews in the course of cultural evolution as a result natural processes of functional differentiation. Rules governing the performance of religious rituals, for example, were distinguished from norms of ordinary human social interaction. The idea that all early human societies had tightly integrated worldviews dominated by religious concerns is, of course, highly speculative. There is little direct evidence that supports it. Perhaps studies of tribal societies by anthropologists during the nineteenth and twentieth centuries lend this idea some indirect evidential support. But the inference from structural features of the worldviews of those tribal societies to structural features of the worldviews of early human societies is problematic. After all, when anthropologists encountered them, the tribal societies they studied had themselves been evolving for a long time.

Moreover, even if something such as this story of the historical origins of morality were true, it would not have important philosophical consequences. It would not establish the conclusion that human beings would never have developed morality if there had been no antecedent religion because a function of large parts of morality is to make possible human cooperation for mutual benefit. People would have encountered problems of cooperation even in the absence of religious beliefs and practices. Given human ingenuity, therefore, it is plausible to suppose that some form of moral belief and practice would have arisen in the course of human history, even if religion had never existed. Nor would history show that the truth of moral beliefs depends on the truth of religious beliefs. In general, it is fallacious to infer from the premise that one belief grew out of another that the truth of the former depends on the truth of the latter. Though modern chemistry grew out of alchemy, it is believed that modern chemistry is true, whereas alchemy is viewed as mostly false.

Morality would depend psychologically on religion if religious beliefs were causally necessary to motivate general compliance with the demands of the moral law. If

human beings are sufficiently selfish, many of them will not behave morally when the moral law requires large sacrifices from them—unless they believe that it is in the long run in their self-interest to do so. The common theistic belief that in the afterlife God rewards those who obey the moral law and punishes those who do not will thus serve to motivate compliance with the demands of moral duty. Maybe this purpose can only be effectively served by a belief that morality has the backing of a system of divine rewards and punishments in the afterlife. If this is the case, people who lack a religious belief of this kind will also lack what it takes to cause or motivate them to live up to the demands of morality when the going gets tough.

However, there are compelling reasons to think that the view of human nature on which this line of thought rests is inaccurate. Living in a social world in which many people lack belief in an afterlife, experience shows that many people are motivated to comply with the most stringent demands of morality even though they lack any belief in a system of divine postmortem rewards and punishments. It was clear to thoughtful people who inhabited social worlds—worlds in which belief in heaven and hell was nearly universal—that belief in divine punishment in the afterlife all too often did not suffice to motivate people who did believe to obey the moral law.

What is more, according to some moral theories, morality requires not only that people comply with the moral law but also that their compliance be motivated by respect for the moral law itself. For example, Kantians hold that actions that are in compliance with the moral law but are motivated by hope for rewards or fear of punishment have no moral worth, even though they are legally correct. In other words, morality demands both that people do their duty and that they do it for duty's sake. They will do the right thing for the wrong reason if their obedience to the moral law is caused by the belief that obedience will be rewarded or the belief that disobedience will be punished. On a view of this sort, religious belief in rewards and punishments in the afterlife constitutes a danger to morality; such belief may tempt people to rely on motivational factors that will deprive their actions of moral value, even when they are the actions prescribed by morality.

### CONCEPTUAL DEPENDENCE

Some philosophers have maintained that concepts of moral deontology contain religious content. In a seminal paper defending a modified divine command account of

wrongness, Robert M. Adams (1987, 1999) proposed a theory in which being contrary to the commands of a loving God is part of the meaning of the term *wrong* in the discourse of some Jewish and Christian theists. And in her famous attack on modern moral philosophy, G. E. M. Anscombe (1981) recommended getting rid of the concepts of moral obligation and moral duty—and the concepts of moral right and wrong—because they belong to an earlier conception of ethics that no longer survives. The earlier conception she had in mind was a law conception. In it, according to Anscombe, the ordinary terms *should*, *needs*, *ought*, and *must* acquired a special sense by being equated in certain contexts with terms such as *is obliged*, *is bound*, or *is required*, in the sense in which one can be obliged or bound—or something be required—legally. She contends that “it is not possible to have such a conception unless you believe in God as a law-giver; like Jews, Stoics and Christians” (Anscombe 1981, p. 30). In the absence of this religious belief, the concepts of moral deontology have no reasonable sense; they are not really intelligible outside a divine law conception of ethics. Modern moral philosophers who lack belief in God would therefore do well to cease using the deontological concepts in their thinking.

Anscombe realizes, of course, that some nonreligious moral theorists will wish to retain a law conception of ethics without a divine legislator. In a Kantian conception of the moral law, for example, practical reason substitutes for God in the role of moral legislator. One's own practical reason engages in self-legislation; it is the authoritative source of moral obligations. Anscombe alleges that the idea of self-legislation is absurd. She remarks: “That legislation can be ‘for oneself’ I reject as absurd: whatever you do ‘for yourself’ may be admirable; but is not legislating” (Anscombe 1981, p. 37). However, she does not offer an argument to support the charge of absurdity. Hence Kantians are in a position to take issue with her cursory dismissal of the idea of moral self-legislation.

A deflationary approach to the deontological concepts provides another nonreligious alternative to the divine law conception. According to the account of this kind proposed by Williams (1983), obligations are not always prescriptively overriding; they do not always beat out ethical considerations of all other kinds. Instead, they are constituted by considerations to which some deliberative priority is granted in order to secure reliability in human social life. High deliberative priority is, in the case of some obligations, responsive to the basic and standing importance of the human interests they serve. Such obligations are negative telling people what not to do. In the

case of positive obligations, high deliberative priority is responsive to the demands imposed by emergencies. Williams thus indicates how it is possible for nonreligious moral theory to salvage at least deflated versions of the concepts of traditional moral deontology.

Anscombe's claim that the main concepts of traditional moral deontology have theistic content is intuitively plausible. However, moral belief and practice seem capable of surviving, almost unchanged, the replacement of such concepts by successors without religious content. And nonreligious moral theorists may even welcome the deflationary features of such a replacement if it is carried out along the lines envisaged by Williams.

### EPISTEMOLOGICAL DEPENDENCE

Many religious believers hold that their moral convictions acquire some positive epistemic status, such as being justified or being warranted, and thereby count as moral knowledge, by virtue of being rooted in religious sources. Among the sources widely acknowledged in theistic religions are divine revelation recorded in sacred texts, divinely inspired prophetic utterances, and the teachers of divinely guided institutions. Frequently such sources purport to reveal divine commands by means of which God promulgates moral obligations. In addition, calls from God to perform particular actions or to enter into religious vocations are taken to be revealed in individual religious experience. Perhaps the most celebrated example in the history of Christianity comes from Augustine's *Confessions*. In retrospect, he took the childish voice he heard saying "Take and read" to be an indirect communication from God, because the biblical reading he did in response served providentially to trigger his conversion to Christianity. Because they hold that these sources are reliable—at least in certain circumstances—theists suppose that their deliverances, when properly interpreted, have positive epistemic status.

Religious diversity furnishes the grounds for an objection to this supposition. Survey the entire religious scene and it becomes evident that there is enormous disagreement among religious people about which sources are reliable, as well as how to interpret the deliverances of these various sources. Consequently, theists disagree among themselves about what God has commanded, and so they disagree about what is morally required or forbidden. Such disagreement undermines the claim that religious sources confer positive epistemic status on their deliverances. Positive epistemic status for one's moral convictions can only be derived from nonreligious sources, because only they can yield agreement. Jeremy

Bentham clearly articulated the epistemic asymmetry implicit in the objection. He remarked: "We may be perfectly sure, indeed, that whatever is right is conformable to the will of God: but so far is that from answering the purpose of showing us what is right, that it is necessary to know first whether a thing is right, in order to know from thence whether it be conformable to the will of God" (Bentham 1948, p. 22). In other words, people do not first come to know, from religious sources, that actions are commanded by God and then, on that basis, come to know that they are morally obligatory. Rather, they first come to know, from nonreligious sources, that actions are morally obligatory and then, on that basis, come to know that they are commanded by God.

Religious disagreement clearly does have a negative impact on the degree to which moral beliefs derive positive epistemic status from religious sources. At least for those who are sufficiently aware of it, religious diversity reduces that degree to a significant extent. After all, moral convictions would acquire a higher degree of positive epistemic status from religious sources if all the sources produced exactly the same outputs. However, nonreligious sources also yield conflicting moral judgments in pluralistic societies that tolerate free inquiry into moral issues. Anyone who is familiar with the history of secular moral theory in the modern era is apt to think it unlikely that agreement on a single moral theory will ever be achieved under conditions of free inquiry. So unless people are prepared to live with extensive moral skepticism, they should be reluctant to think that moral beliefs derive no positive epistemic status at all from religious sources merely because those sources yield conflicting deliverances.

Few people who live in religiously pluralistic societies rely exclusively on religious sources for epistemic support of their moral beliefs. Most people think the moral beliefs they form when responding intuitively to their experiences or to works of imaginative literature—or those beliefs acquired from interaction with parents and peers outside of religious contexts—often have positive epistemic status bestowed on them by nonreligious sources of these kinds. Even the religious people who inhabit such societies typically find themselves with moral convictions that stem from a plurality of sources, some religious and others nonreligious. However, unless the religious worldviews that serve to accredit their religious sources are disqualified for rational acceptance—which would be difficult to establish—religious people seem to be entitled to trust those religious sources and to regard them as conferring positive epistemic status on

their deliverances. Hence the moral convictions of religious believers apparently can, in principle, derive positive epistemic status from both religious and nonreligious sources. Bentham's view is therefore one-sided. While religious believers in pluralistic societies may acquire knowledge of what God commands by first coming to know their obligations, they may also acquire knowledge of their obligations by first coming to know what God commands. At least some of the moral convictions of such people can be epistemologically dependent on their religious beliefs and yet possess positive epistemic status. Or, at any rate, this view is more plausible than Bentham's if moral and religious skepticism is ruled out.

### METAPHYSICAL DEPENDENCE

Beginning in the last third of the twentieth century, interesting ideas about how morality might depend metaphysically on God were developed and defended in the work of proponents of divine command theories of morality. In an influential paper offering suggestions to divine command theorists, William P. Alston (1990) proposed that axiology and deontology depend on God in different ways. In the axiological domain, in Alston's view, God is the paradigm or supreme standard of goodness. An analogy to the situation helps to clarify Alston's suggestion. He maintained that the meter could be defined in terms of a certain metal bar kept in Paris. What then made a particular table a meter in length was its conformity to a certain existing individual. Similarly, according to Alston, "what ultimately makes an act of love a good thing is not its conformity to some general principle but its conformity to, or approximation to, God, Who is both the ultimate source of the existence of things and the supreme standard by reference to which they are to be assessed" (Alston 1990, p. 320). There is, to be sure, a disanalogy as well. While it is arbitrary which particular physical object was chosen to be the standard meter, Alston does not suppose that it is similarly arbitrary whether God or someone else serves as the standard of goodness. Thus understood, moral axiology depends metaphysically on the nature and character of God. By contrast, within the domain of deontology, moral obligations and moral wrongness depend metaphysically on God's commands, and ultimately on the divine volitions expressed by those commands.

Alston's suggestions have been developed into a framework for ethics by Robert M. Adams. According to his theistic Platonism, God plays the role that the Form of the Good plays in Plato's metaphysics. God is the Good Itself, the standard of goodness; and other things are good

by virtue of resembling or being images of God in various ways. Modifying again his modified divine command theory of wrongness, Adams has claimed that wrongness bears the metaphysical relation of property-identity to contrariety to the commands of a loving God. He asserts: "My new divine command theory of the nature of ethical wrongness, then, is that ethical wrongness *is* (i.e., is identical with) the property of being contrary to the commands of a loving God" (Adams 1987, p. 139). And in presenting his framework for ethics, Adams sometimes says that an action's being obligatory consists in its being commanded by a loving God and that an action's being wrong consists in its being contrary to the commands of a loving God. The fundamental principle of obligation of a theory of this kind asserts that actions are obligatory if and only if, and solely because, they are commanded by a loving God. Its fundamental principle of wrongness claims that actions are wrong if and only if, and solely because, they are forbidden by a loving God. The metaphysical dependency of moral deontology on God is expressed in such principles by their requirement that actions are obligatory or wrong just because a loving God commands or prohibits them.

Of course, many philosophers have mounted objections to divine command theories of morality. Perhaps the most famous objection alleges that divine command theories render moral deontology arbitrary because God could have commanded absolutely anything. Thus, for example, God could have made cruelty for its own sake obligatory simply by commanding it. A defense against this allegation is available within the framework proposed by Alston and developed by Adams. God's nature and character, which constitute the standard of goodness, constrain what God can command. Though they may well leave some room for discretion in what God commands, God cannot command absolutely anything. If God is essentially loving and so could not be otherwise, it is impossible for God to command cruelty for its own sake. Hence, according to a divine command theory of this sort, it is likewise impossible for cruelty for its own sake to be obligatory.

Divine command theories have been defended against many other objections in work by Philip L. Quinn (1978) and Edward R. Wierenga (1989). As a result, it seems that these theories are good candidates for adoption by theists. If the larger theistic worldviews in which divine command theories are embedded are themselves rationally acceptable, an account of the metaphysics of morals, according to which morality depends on God, is a live option in moral theory.

## CONFLICT THREATENED AND RESISTED

The arbitrariness objection to divine command theories suggests a threat that religion may—in some cases—pose to morality. It is the possibility of a religious obligation, imposed by divine command, coming into conflict with moral duties. The possibility is ominous because the historical record is full of crusades, inquisitions, and terrorist acts perpetrated in the name of theistic religions. Those who have done such things have often sincerely believed that they act in obedience to God's will. Within Jewish and Christian traditions, reflection on this possibility frequently focuses on the Hebrew Bible's story of the *akedah*, the binding of Isaac, narrated in Genesis 22. According to the story, God commands Abraham to offer his innocent son, Isaac, as a sacrifice, and Abraham shows that he is willing to perform this terrible deed of human sacrifice. As it turns out, an angel tells Abraham that he is permitted to substitute a ram for Isaac as the sacrificial victim, but the substitution is permitted precisely because Abraham has demonstrated to God his willingness not to withhold Isaac from being killed as a sacrifice.

Johannes de Silentio, the pseudonymous author of Søren Kierkegaard's *Fear and Trembling*, argues that the story of the *akedah* reveals a teleological suspension of the ethical. De Silentio conceives of the ethical in broadly Hegelian terms. People have *prima facie* duties to social groups of various size. If a duty to a smaller group conflicts with a duty to a larger group, the duty to the larger group is more stringent than—and hence overrides—the duty to the smaller group.

Thus, for example, Agamemnon's familial duty not to sacrifice his innocent daughter, Iphigenia, is overridden by his political duty to lead the Greek expedition to Troy. He is a tragic hero because he sacrifices Iphigenia. However, he remains within the ethical in doing so because he does so in order to fulfill his overriding political duty. Abraham is not a tragic hero. When he consents to sacrifice Isaac, he does not do so in order to fulfill some more stringent duty to a larger social group. Were he to carry out the sacrifice, he would be violating a duty that has not been overridden within the ethical. Yet Abraham lies under an absolute religious obligation to obey God. De Silentio regards Abraham's situation as a paradox that cannot be solved by mediation. He claims: "During the time before the result, either Abraham was a murderer every minute or we stand before a paradox that is higher than all mediations" (Kierkegaard 1983, p. 66). In other words, from the time he consents to sacrifice Isaac, Abraham is a murderer in his heart unless the ethical is sus-

pending from the outside in his case. But Abraham, whom de Silentio acknowledges to be the Father of Faith, is never a murderer. Therefore the divine command to Abraham must produce a suspension of the ethical.

Many theists do not wish to accept such a radical interpretation of the *akedah*. Kant is a notable example. In an often cited footnote in *The Conflict of the Faculties*, he insists: "Abraham should have replied to this supposedly divine voice: 'That I ought not to kill my good son is quite certain. But that you, this apparition, are God—of that I am not certain, and never can be, not even if this voice rings down to me from (visible) heaven'" (Kant 1996, p. 283). Kant's strategy of resistance to radical readings of the *akedah* carries with it an epistemological price. No matter how impressive the sound effects in the sky may be, they cannot confer on the claim that the voice commanding Abraham to kill Isaac actually came from God the exalted positive epistemic status of certainty. More generally, religious sources cannot confer epistemic certainty on claims about what God has commanded that conflict with epistemically certain moral judgments. On this Kantian view, therefore, there are limits on the extent to which claims about what God commands or wills can derive positive epistemic status from religious sources. No doubt this is a price many theists will be happy to pay in order to rule out certain sorts of conflict between their religious beliefs and the moral beliefs to which they are most deeply committed.

**See also** Atheism; Authority; Bentham, Jeremy; Deontological Ethics; Enlightenment; Ethics, History of; Hobbes, Thomas; Kant, Immanuel; Locke, John; Mill, John Stuart; Philosophy of Education, History of; Philosophy of Law, History of; Philosophy of Religion, History of; Rashdall, Hastings; Teleological Ethics.

### Bibliography

- Adams, Robert M. *Finite and Infinite Goods: A Framework for Ethics*. New York and Oxford: Oxford University Press, 1999. Develops a framework in which a divine command deontology is embedded in an axiological theistic Platonism.
- Adams, Robert M. *The Virtue of Faith and Other Essays in Philosophical Theology*. New York and Oxford: Oxford University Press, 1987. Reprints the 1973 paper in which his modified divine command theory of wrongness was first proposed and the 1979 paper in which that theory was further modified.
- Alston, William P. "Some Suggestions for Divine Command Theorists." In *Christian Theism*, edited by Michael D. Beaty. Notre Dame, IN, and London: University of Notre Dame Press, 1990. Suggests combining an axiology in which God is



- the standard of goodness with a divine command deontology.
- Anscombe, G. E. M. *Ethics, Religion and Politics*. Minneapolis: University of Minnesota Press, 1981. Reprints the 1958 paper that contains her famous attack on modern moral philosophy.
- Bentham, Jeremy. *An Introduction to the Principles of Morals and Legislation* (1789). New York: Hafner, 1948. Argues that morality is not epistemologically dependent on religion.
- Geertz, Clifford. "Religion as a Cultural System." In *Anthropological Approaches to the Study of Religion*, edited by Michael Banton. New York: Praeger, 1966. Presents a definition of religion according to which the postulation of superhuman beings is not a necessary condition for being a religion.
- Hick, John. *An Interpretation of Religion: Human Responses to the Transcendent*. New Haven, CT, and London: Yale University Press, 1989. Endorses a family-resemblance account of the concept of religion.
- Kant, Immanuel. *The Conflict of the Faculties* (1798). In Immanuel Kant, *Religion and Rational Theology*, edited and translated by Allen W. Wood and George di Giovanni. Cambridge, U.K.: Cambridge University Press, 1996. Comments on the *akedah* in the course of a discussion of the relations of theology and philosophy.
- Kant, Immanuel. *Religion within the Boundaries of Mere Reason* (1793). In his *Religion and Rational Theology*, edited and translated by Allen W. Wood and George di Giovanni. Cambridge, U.K.: Cambridge University Press, 1996. Develops his account of moral religion, which contains reflections on the *akedah*.
- Kierkegaard, Søren. *Fear and Trembling* (1843). In his *Fear and Trembling/Repetition*, edited and translated by Howard V. Hong and Edna H. Hong. Princeton, NJ: Princeton University Press, 1983. Interprets the *akedah* as involving a teleological suspension of the ethical.
- Outka, Gene, and John P. Reeder eds. *Prospects for a Common Morality*. Princeton, NJ: Princeton University Press, 1993. Collects papers by scholars in philosophy and religious studies that address the question of whether religion can contribute to a common morality for a pluralistic world.
- Outka, Gene, and John P. Reeder, eds. *Religion and Morality*. Garden City, NY: Doubleday, 1973. Collects papers by scholars in philosophy and religious studies on the relations of religion and morality.
- Quinn, Philip L. *Divine Commands and Moral Requirements*. Oxford, U.K.: Clarendon Press, 1978. Defends divine command theories against a variety of objections.
- Spiro, Melford E. "Religion: Problems of Definition and Explanation." In *Anthropological Approaches to the Study of Religion*, edited by Michael Banton. New York: Praeger, 1966. Presents a definition of religion according to which the postulation of superhuman beings is a necessary condition for being a religion.
- Wierenga, Edward R. *The Nature of God: An Inquiry into Divine Attributes*. Ithaca, NY, and London: Cornell University Press, 1989. Formulates and defends a divine command theory in which deontology depends metaphysically on divine agent causation.
- Williams, Bernard. *Ethics and the Limits of Philosophy*. Cambridge, MA: Harvard University Press, 1985.
- Distinguishes morality from ethics and offers a deflationary account of moral obligation.
- Philip L. Quinn (2005)*

## RELIGION AND POLITICS

Is it morally appropriate for citizens in a liberal democracy like the United States to support or oppose public policies solely for religious reasons? Although regularly serving as grist for the mill of political theorists, that question is not the familiar fare of ordinary political discussion. It's not a question about, or at least directly about, which laws our government ought to enforce. We're all too familiar with such questions—about the moral propriety and practical wisdom of abolishing the death penalty, legalizing abortion, declaring war on Afghanistan, and so on. Rather, it is a question about the kinds of *justifications* citizens should or should not have for their political commitments.

The most common position on this issue calls for a general constraint on the political use of religious reasons. Proponents of this constraint argue that citizens must support public policies for secular reasons and therefore that they morally ought to restrain themselves from supporting public policies solely for religious reasons. So, for example, they argue that a citizen who lacks any secular reason to criminalize abortion or discourage homosexuality ought to refrain from supporting any such policy.

This entry presents the main lines of argument for and against this view that citizens should obey the *doctrine of restraint*: that they ought to restrain themselves from supporting or opposing public policies solely for religious reasons.

### LIBERAL DEMOCRACY, RELIGIOUS PLURALISM, AND THE DOCTRINE OF RESTRAINT

In order to understand why a given philosophical commitment is significant, it's helpful to identify the problem that that commitment is supposed to solve. This is true of the doctrine of restraint: It is significant because it claims to solve a problem that naturally arises from the institutionalization of a liberal democracy's deepest normative commitments. What is that problem?

At the very least, a liberal democracy is a form of government that affirms and protects a citizen's rights—to private property, to freedom of association, to freedom

of conscience, and so on. The most fundamental of these rights, both morally and historically, is the right to religious freedom: Citizens are entitled to decide which religious creed or practice, if any, they wish to pursue. So a liberal democracy just is a kind of political system that provides citizens with considerable leeway to decide for themselves what they are to believe regarding religious matters.

Protection of the right to religious freedom has far-reaching social consequences, the most important of which is religious pluralism: A society that assiduously protects each citizen's right to religious freedom will find its citizens disagreeing among themselves as to which religion is true, how to please God, and so on. What explains this close connection between religious freedom and religious pluralism? Because our rational capacities are not powerful enough to produce widely convincing proofs (or refutations) of religious truth claims, even the flawless employment of our rational capacities will lead to disagreements about such matters. From this claim about the limited powers of our rational capacities, it follows that citizens who are free to decide which religious tradition to affirm will embrace a diversity of religious traditions. So, then, since well-functioning liberal democracies effectively protect the right to religious freedom, and since the effective protection of religious freedom results in a citizenry that is rationally committed to divergent religious traditions, it follows that well-functioning liberal democracies will be characterized by a citizenry that is rationally committed to divergent religious traditions.

This fact of religious pluralism raises a question of enormous moral and practical import: How are the multifariously committed citizens of a liberal democracy to make collective decisions about the laws with which each citizen must comply? For advocates of the doctrine of restraint, the pervasive pluralism of a liberal democracy renders obedience to the doctrine of restraint imperative: The morally appropriate response of citizens to religious pluralism is to refrain from resolving public matters solely for sectarian, and therefore for religious, reasons.

### THE DOCTRINE OF RESTRAINT

Although agreed on a core prohibition of exclusively religious support for public policies, advocates of the doctrine of restraint diverge in their formulations of this doctrine. Some have argued that it constrains all political actors, including citizens, legislators, judges, and other public officials; others limit its scope to public officials. Some have argued that the doctrine enjoins restraint with respect to all public policies, whereas others have limited

its scope to coercive public policies; others further delimit its scope to laws of fundamental and structural importance. Some have argued that the doctrine of restraint applies only to public political advocacy, whereas others contend that it should also apply to political decision-making the decision as to whether to vote for some candidate, for example.

A more important variation is between the inclusive formulation of the doctrine of restraint on which this entry focuses and a more demanding but less plausible cousin with which the inclusive version is sometimes confused. The inclusive version of the doctrine of restraint has been most effectively advocated by Robert Audi (2000), who insists that citizens should *include* secular arguments in their political practice, not that they should *exclude* religious reasons. According to Audi, citizens may support only those public policies for which they have secular reasons they regard as sufficiently weighty that they would continue to support the relevant policies absent corroborating religious reasons. But as long as citizens have, and are motivated by, adequate secular reasons for a given public policy, they are free to appeal to religious reasons as well. So inclusivists such as Audi advocate a kind of *limited* privatization: Citizens are free to rely on religious reasons for public policies so long as they have correlative secular reasons but must refrain from supporting public policies for which they have only religious reasons

Other advocates of the doctrine of restraint have demanded a complete privatization of religious reasons. This *exclusive* version of the doctrine of restraint, advocated by Richard Rorty (1994), demands that citizens refrain entirely from relying on religious reasons when supporting or opposing public policies. Its advocates expect that the real business of politics will be conducted exclusively on the basis of secular argument. But this complete privatization of religious reasons is gratuitously exclusionary: There is no good reason to stigmatize citizens who support a given public policy on religious grounds if those citizens also have and are sufficiently motivated by plausible secular reasons. Any serious evaluation of the doctrine of restraint, therefore, will pay due regard to the more moderate, inclusive version articulated by Audi.

### AGAINST THE DOCTRINE OF RESTRAINT

Critics of the doctrine of restraint reject even its inclusive formulation. According to Eberle, (2002), Perry (2003), and Audi and Wolterstorff (1997), citizens need not be

morally criticizable in any respect for supporting public policies for religious reasons—even solely for religious reasons. These critics of the doctrine of restraint need not and do not typically license an “anything goes” approach. They may advocate *substantive* constraints on the reasons citizens have for their favored public policies—for example, they might argue that citizens ought not appeal to reasons that deny the dignity of their fellow citizens. They may also advocate *epistemic* constraints on the manner in which citizens support public policies, arguing, for example, that citizens should engage in critical reflection on their reasons. They may even argue that citizens should *try* to articulate secular reasons for their favored political commitments: Citizens ought to do what’s within their power to speak to their fellow citizens in ways that their fellow citizens can take seriously and so ought to do what’s in their power to articulate reasons that speak to their secular compatriots—presumably these will be secular reasons. Critics of the doctrine of restraint argue, however, that none of these constraints provide an adequate basis for a general constraint on religious reasons; so long as a citizen satisfies the appropriate substantive and epistemic constraints, and so long as he or she genuinely searches for a plausible secular rationale, then a citizen has no good reason not to support a given public policy solely for religious reasons.

These critics argue that advocates of restraint must discharge a heavy burden of proof: Absent sufficiently powerful reasons in favor of the doctrine of restraint, citizens may refuse without compunction to comply with that doctrine. The argument for this distribution of the burden of proof is short, direct, and powerful. We surely want and expect citizens to treat their compatriots as conscience dictates: A citizen ought to support or oppose public policies on the basis of what he or she sincerely and responsibly believes to be the just and decent thing to do. And sometimes what a citizen sincerely takes to be the just and decent thing to do will depend solely on religious beliefs. And in that case, the heavy presumption in favor of acting in accord with conscience translates into a heavy presumption permitting a citizen to decide, solely on religious grounds, to support or oppose some public policy. Consider, for example, a Christian pacifist who, after sober and competent reflection on the morality of war, concludes that the life and teachings Jesus Christ forbid the lethal use of force. In this case, our conviction that citizens should support those public policies that they actually believe to be morally correct should lead us to expect—indeed, encourage—Christian pacifists to oppose war, even though they have an exclusively religious rationale for that policy.

So critics of the doctrine of restraint will appeal to the very great good of a citizen’s acting in accord with her conscience to establish a heavy presumption in favor of the moral propriety of that citizen’s making political decisions solely on religious grounds. But that there is a presumption against the doctrine of restraint by no means implies that that doctrine is false. After all, presumptions can be overridden, and the burden of proof can be met—so long as advocates of the doctrine of restraint can marshal sufficiently powerful arguments.

### THE ARGUMENT FROM RESPECT

Some advocates of the doctrine of restraint have argued that citizens should obey the doctrine of restraint out of respect for their compatriots. When a citizen supports a public policy, she is complicit in authorizing the government to coerce citizens. But her compatriots aren’t mere playthings who may be forced to satisfy her whims; rather, they’re rational persons who are fully capable and desirous of deciding for themselves how they will live their lives on the basis of reasons *they* find acceptable. And so if she is to respect her compatriots as persons, she must be committed to providing them with reasons that they find, or at least can find, acceptable. That requires a search for some common ground, premises that one might *share* with one’s compatriots. Given the pervasive religious pluralism of a well-functioning democracy, this common ground will most likely be secular, not religious in content. On this view, advocated by Charles Larmore (1987), it is respect for the dignity and autonomy of our fellow citizens that requires us to abide by the doctrine of restraint.

The argument from respect is both popular and controversial. Critics have expressed doubt that reliance on religious reasons—even exclusive reliance—necessarily involves disrespect for other persons. Some argue that it is unclear why any disrespect can be imputed to citizens who affirm their compatriots’ dignity, who are willing to engage in critical analysis of their favored public policies, and who provide their fellow citizens with sincerely held and carefully elaborated reasons that are, nevertheless, based on religious doctrines. Consider again the Christian pacifist: There is every reason to believe that her opposition to war is based on the kind of moral commitment that putatively underlies the doctrine of restraint: respect for *all* other persons. The argument from respect seems an entirely unpromising rationale for requiring a Christian Pacifist to exercise restraint—thereby casting doubt on the doctrine of restraint insofar as it constrains on religious reasons generally.

## THE ARGUMENT FROM RELIGIOUS WARFARE

Religious wars have played a defining role in the history of liberal democracies; the commitment to religious freedom was formulated and defended in reaction to a century and a half of wars fought to “resolve” religious disagreements. The specter of religious warfare lingers on, and in some cases, that wariness motivates the argument from religious warfare.

Here is one way to formulate that argument. Religious wars are morally abhorrent: Military conflicts guided by religious aims are purely destructive, extraordinarily vicious, and utterly without redeeming value. If large numbers of citizens rely solely on religious reasons to direct state coercion, there is a glaring temptation to enlist the power of the state to force conversion and persecute heretics, thereby provoking armed conflict; hence only a policy of religious restraint can ward off the specter of sectarian bloodshed. In short, that citizens firmly commit to supporting only those public policies for which they have an adequate secular rationale is a crucial bulwark protecting us from confessional conflict.

It is, however, reasonable to deny that there is a *realistic* prospect that segments of the population of the United States will enter into armed conflict over religious matters. Religious warfare is not a realistic prospect in the contemporary United States because we have learned how to prevent it and have taken the appropriate measures: The proper preventive for religiously generated strife is constitutional and cultural, viz., effective protection of religious freedom on the part of the government and commitment to religious freedom on the part of citizens. This point has direct implications for the idea that obedience to the doctrine of restraint is necessary to prevent religious war. For it implies that what’s essential in preventing religious war is that citizens are fully committed to religious freedom, not that they refrain from making use of that right to support public policies solely for religious reasons. So long as citizens are firmly committed to religious freedom, their willingness to support public policies solely for religious reasons has no realistic prospect of engendering religious warfare.

## THE ARGUMENT FROM PUBLIC DISCOURSE

A third argument for the doctrine of restraint, advocated by Daniel Conkle (1993–1994) hinges on the following two claims: (1) that healthy public discussion of public policies is a great moral and political good and (2) that

that good would be threatened by the refusal of large numbers of citizens to abide by the doctrine of restraint. On this view citizens should not support public policies without reflecting on those policies with their compatriots; as an implicit acknowledgment of human fallibility, the pursuit of such political discourse invites our compatriots to challenge our mistaken assumptions and inherited prejudices. Moreover, a commitment to public discourse about public policies affords those who hold a minority view the opportunity to convince other citizens of good will that their minority position is in fact correct. Hence this kind of public discourse is advanced as an important moral good.

In order to secure that good, citizens must abide by a number of constraints, especially that which requires citizens to support public policies on the basis of reasons open to rational evaluation and debate. Religious reasons, by contrast, are not subject to rational analysis and thus require a nonrational, subjective act of faith that can only be experienced, not rationally analyzed or debated. On this view compliance with the doctrine of restraint is a prerequisite for healthy discourse about public policies.

Critics have pointed to a number of problems in the argument from public discourse. Insofar as advocates of the doctrine of restraint depend heavily on the argument from public discourse, they seem to rely on controversial claims about the epistemic status of religious reasons. If religious reasons are not amenable to rational criticism by others, then it follows that religious reasons lack what many regard as an important epistemic desideratum. But this demotion of the epistemic status of religious reasons likely to trouble religious citizens. As this entry noted at the outset, the primary significance of the doctrine of restraint is that it putatively provides a morally attractive guideline for a pluralistically committed citizenry to follow when supporting public policies. This implies that the doctrine of restraint should be acceptable not just to secular citizens, but to all citizens and therefore to the religious citizens who are expected to comply with that doctrine. But many religious believers are likely to regard the epistemic assessment of religious claims that underpins the argument from public discourse as thoroughly objectionable; after all, it’s dubious that we should place our trust in claims, whatever their content, that aren’t amenable to rational criticism by others. So the argument from public discourse recommends that religious believers exercise restraint, but on grounds that many religious believers will find deeply objectionable. It seems likely, then, that the best argument advocates of the doctrine of restraint can muster will be anathema to the very citizens

who are expected to comply with that doctrine, thus emptying the doctrine of restraint of its primary significance.

## CONCLUSION

The literature on the proper role of religious reasons in liberal politics is voluminous. And so as one might expect, the preceding discussion is far from definitive (or exhaustive for that matter). But this should hardly be surprising: The problem to which the doctrine of restraint responds rests on a pluralistic social reality that results from the successful implementation of a liberal democracy's defining commitments. That social reality is here to stay, as are the problems that it engenders, and so reflective people will continue to advocate for and criticize proposed solutions to those problems. The doctrine of restraint, and its critics, will be with us for the foreseeable future

**See also** Democracy; Liberalism; Philosophy of Religion, History of; Political Philosophy, History of; Rawls, John; Social and Political Philosophy.

## Bibliography

- Audi, Robert. *Religious Commitment and Secular Reason*. Cambridge, U.K.: Cambridge University Press, 2000.
- Audi, Robert, and Nicholas Wolterstorff. *Religion in the Public Square: The Place of Religious Convictions in Political Debate*. Lanham, MD: Rowman and Littlefield, 1997.
- Carter, Stephen. *The Culture of Disbelief*. New York: Basic Books, 1993.
- Conkle, Daniel. "Different Religions, Different Politics." *Journal of Law and Religion* (1993–1994).
- Eberle, Christopher. *Religious Convictions in Liberal Politics*. Cambridge, U.K.: Cambridge University Press, 2002.
- Greenawalt, Kent. *Private Consciences and Public Reasons*. Oxford: Oxford University Press, 1995.
- Greenawalt, Kent. *Religious Convictions and Politics Choice*. Oxford: Oxford University Press, 1988.
- Larmore, Charles. *Patterns of Moral Complexity*. Cambridge, U.K.: Cambridge University Press, 1987.
- Perry, Michael. *Love and Power: The Role of Religion and Morality in American Politics*. Oxford: Oxford University Press, 1991.
- Perry, Michael. *Religion in Politics: Constitutional and Moral Perspectives*. Oxford: Oxford University Press, 1997.
- Perry, Michael. *Under God? Religious Faith and Liberal Democracy*. Cambridge, U.K.: Cambridge University Press, 2003.
- Rawls, John. *Political Liberalism*. New York: Columbia University Press, 1993.
- Rorty, Richard. "Religion as a Conversation-Stopper." *Common Knowledge*. (1994).
- Weithman, Paul ed. *Religion and Contemporary Liberalism*. Notre Dame, IN: University of Notre Dame Press, 1997.

Weithman, Paul ed. *Religion and the Obligations of Citizenship*. Cambridge, U.K.: Cambridge University Press, 2002.

Quinn, Philip L. "Political Liberalism and their Exclusions of the Religious." In *Religion and Contemporary Liberalism*, edited by P. J. Weithman, 138–161. Notre Dame, IN: Notre Dame Press, 1977.

*Christopher J. Eberle (2005)*

## RELIGION AND THE BIOLOGICAL SCIENCES

### HISTORY

Plato and Aristotle recognized that understanding nature demands reference to factors—what Aristotle called "final causes"—that in some sense anticipate what will or should happen. In the *Timaeus*, Plato wrote, "From the combination of sinew, skin, and bone, in the structure of the finger, there arises a triple compound, which, when dried up, takes the form of one hard skin partaking of all three natures, and was fabricated by these second causes, but designed by mind, which is the principle cause with an eye to the future." He continued, "For our creators well knew ... that many animals would require the use of nails for many purposes; wherefore they fashioned in men at their first creation the rudiments of nails. For this purpose and for these reasons they caused skin, hair, and nails to grow at the extremities of the limbs" (*Timaeus*, 76d–e).

Such adaptations, organic features that demand a final-cause understanding, are the basis for (what was to prove) a very popular and longstanding proof of God's existence. The forward-looking aspect of adaptations comes from the fact that they seem as if they were designed. They are like artifacts. Why? Quite simply because adaptations are artifacts—the artifacts of a deity. Just as a couch has a couch designer, so the hand and the eye must have a hand and eye designer. There is no necessary implication that there is just one designer, or that it has the attributes of the Judeo-Christian God—eternal, all powerful, all loving, creator of all from nothing—but this Greek argument (known as the "argument from design") was taken over by the great Christian philosophers and theologians, and became one of the main supports of the route to God through reason (natural theology).

This argument continued to enjoy great popularity and force right into the nineteenth century. Archdeacon William Paley in his book *Natural Theology* (1802) pro-

moted the argument: The eye is like a telescope; telescopes have telescope makers; therefore eyes must have eye makers—what one might call the Great Optician in the sky. By this time, however, the pendulum was starting to swing the other way, with biology giving theists cause for concern. The eighteenth century saw the rise of evolutionary speculations—hypotheses that organisms are the end results of long, slow, natural processes of development from very different and much simpler forms. At the most obvious level, evolutionary ideas challenge the Genesis story of creation. But though this was certainly a stumbling block for many, believers have long had resources to deal with problems caused by literal interpretations of the Bible.

Far more threatening to the theist was the connection between organic evolution and the doctrine of intellectual or cultural progress. As humans supposedly have risen up from ignorance and poverty in the cultural world to the sophisticated state in which we humans now find ourselves, so in the world of organisms, primitive forms have developed into humans. Cultural development points to biological evolution, which in turn reinforces cultural development. To quote an early evolutionist, Erasmus Darwin (a grandfather of Charles):

Imperious man, who rules the bestial crowd,  
Of language, reason, and reflection proud,  
With brow erect who scorns this earthy sod,  
And styles himself the image of his God;  
Arose from rudiments of form and sense,  
An embryo point, or microscopic ens!  
(1803, 1, 295–314)

All of this progressivism was a direct challenge to the Christian notion of Providence. For the believer, because of Adam's sin, we are in a fallen state. To earn us salvation in this fallen state, God intervened in his creation, choosing freely to die on the cross. This means that our happiness comes not from our merits, but simply as the result of God's forgiveness and grace. Progress challenges this. It carries the central message that improvement is possible and due entirely to human intentions and labors. Success comes from our own efforts, not from those of others—including God. As part of the picture of progress, evolution was rightly seen as challenging conventional religious verities.

Although popular in some quarters, evolution was always somewhat of a pseudoscience. As Immanuel Kant pointed out in his third critique, *The Critique of Judgment*, there are difficulties with final causes. Such a complex, apparently intentional entity as the eye simply could have come about through blind law. Charles Darwin

addressed this issue in *On the Origin of Species*, published in 1859. Committed to evolution, Darwin sought a cause that would speak to adaptation. This he found in the mechanism of natural selection. More organisms are born than can survive and reproduce. This brings on a struggle for existence. Organisms tend to vary naturally, and the winners in the struggle (the fit) have features not possessed by the losers (the unfit). Moreover, these features tend to be deciding factors in whether an organism is successful or unsuccessful. Hence, equivalent to the selection practiced by animal and plant breeders, there is a natural selection, where the winners pass on their favorable features. Over time this leads to full-blown evolution, a key feature of which is the development and perfection of adaptations.

Although he himself was never an atheist—at the time of writing the *On the Origin of Species* he was a deist and later turned to agnosticism—Darwin apparently drove a stake through the heart of the argument from design. The eye resulted from blind, unguided processes through natural selection. There is no need to invoke a designer. In the words of the contemporary English biologist Richard Dawkins (1986), only after Darwin was it possible to be “an intellectually fulfilled atheist.” As expected, not everyone agrees that such a conclusion follows. Below are the different positions taken on the relation of biology and religion in the post-Darwinian era.

## SEPARATION

One strategy is to separate science and religion, specifically, biology and Christianity. This means that biology cannot support religion, but then again neither can it refute it. A common suggestion is that biology can tell us how things occur—that humans came from apelike creatures, for example—but it cannot tell us why things occur—why there should be creatures with the conscious ability to tell good from evil. The great English theologian John Henry Newman, an Anglican convert to Catholicism, had no trouble at all with evolution. It was simply not something that bore on his faith. “I believe in design because I believe in God; not in a God because I see design.” He continued, “Design teaches me power, skill and goodness—not sanctity, not mercy, not a future judgment, which three are of the essence of religion” (Newman 1973, 97).

This kind of reversal of the argument—design because of God, rather than God because of design—found much favor in the twentieth century, particularly in circles influenced by Karl Barth, another major critic of natural theology. In the opinion of such thinkers, often

labeled “neo-orthodox,” evolution is true. But this does not prove anything affecting religion. In the language of the German theologian Wolfhart Pannenberg (1993), we must strive for a “theology of nature,” where the beauties of the living world enrich our faith, rather than a “natural theology,” where the living world is used as a substitute for faith. Thus, Dawkins is wrong not so much in thinking that one can be an intellectually fulfilled atheist, but in thinking that this is the end of the journey. It is the beginning. Darwin shows that there can be no proofs, and that is where faith begins.

## INTERACTION

Not every post-Darwinian thinker has been so negative about natural theology. Many think that Darwin’s work is the spur to find a new natural theology, a natural theology that accepts evolution and works with it rather than against it. Instead of rejecting progress, Christians should take it on board in some fashion, arguing that we humans should work with God to achieve our salvation. The rise of organisms, from slime to humans, “from monad to man,” as it was traditionally put, is proof that not all is random and without purpose. It shows that God is working out his plan, and also that we are obligated to work with him.

The thinker who tried most fully to work out a theology that stayed true to conventional Christian belief and yet made the upward progressive message of evolutionism central was the French Jesuit and paleontologist Pierre Teilhard de Chardin. In his masterwork *The Phenomenon of Man*, Teilhard saw life evolving upward through the realm of life (the biosphere), to the realm of humans and consciousness (the noosphere), and then even further onward and upward to the Omega Point, which in some way he identified with the Godhead, with Jesus Christ. “An ever-ascending curve, the points of transformation of which are never repeated; a constantly rising tide below the rhythmic tides of the ages—it is on this essential curve, it is in relation to this advancing level of the waters, that the phenomenon of life, as I see things, must be situated” (p. 101).

One major problem with this whole approach is less whether the post-Darwinian Christian should accept the doctrine of progress than whether the post-Darwinian evolutionist should accept such a doctrine. If natural selection is true, then change is much relativized. Which species are fit? Not necessarily those at the top of an absolute scale. Intelligence might seem a good thing, but it has major costs, not the least of which is a constant supply of quality protein. In many circumstances, stupidity

and strength might be a better biological strategy. Many evolutionists now reject progress entirely. The late Stephan Jay Gould (1989), paleontologist and science writer, argued that there is no genuine progress, and certainly no guarantee that if the tape of life were replayed, humans would inevitably emerge.

This is not the last word. Darwin himself believed in progress and thought that natural selection gives rise to what biologists of 2005 label “arms races,” where one line of organisms competes and improves adaptations against the threat of other lines. Intelligence is an end result. Darwin has his supporters in the early twenty-first century, notably the English paleontologist Simon Conway Morris (2003), who argues that selection leads steadily to the conquering of one major ecological niche after another. Consciousness is the prize at the top, waiting to be grasped, and if not by humans, then by some other contender with outstretched paw.

## DARWINIAN OPPOSITION TO THEISM

Dawkins is an atheist. He thinks that Darwinian evolution is hardly neutral. Although the argument from evil—that the bad things of this world are incompatible with an all-loving, all-powerful god—is not new with Darwin, his theory focuses on evil and makes it a central part of the evolutionary story. For Dawkins and others, this is confirmation that the Christian God does not exist, that other forms of deity are not worth entertaining, and hence that life has no meaning, that it just is. “In a universe of blind physical forces and genetic replication, some people are going to get hurt, other people are going to get lucky, and you won’t find any rhyme or reason in it, nor any justice. The universe we observe has precisely the properties we should expect if there is, at bottom, no design, no purpose, no evil and no good, nothing but blind, pitiless indifference” (Dawkins 1995, 133).

Theists have standard counters to the problem of evil (Ruse 2001). Some theists separate moral evil (the extermination of Jews at Auschwitz) from physical evil (cancer). In the case of moral evil, it is better that humans have free will, even though they will do wrong, these theists argue, than that humans have no genuine choices at all. This may or may not be an adequate response, but if one argues for the philosophical position known as compatibilism—the position that freedom and natural law are not contradictory—then an evolutionist could in principle support this defense. The fact that we humans are the product of biological law and still subject to it does not in itself deny some dimension of freedom and ability to act on our own choices.

In the case of physical evil, recourse is often made to an argument of Gottfried Wilhelm Leibniz (1646–1716), namely that such evil is an unfortunate but unpreventable consequence of a world governed by natural law. Here too the evolutionist has a defense. Somewhat paradoxically, Dawkins himself supports this counter, for he argues that if organisms were created naturally, then adaptive complexity could have been achieved only through the action of natural selection. “The Darwinian Law ... may be as universal as the great laws of physics” (Dawkins 1983, 423). One might still argue that given the consequent pain, it was a pity that God created at all, but this is a different claim totally independent of evolution. From the viewpoint of biology, if God did create and did so through natural law—and there may be good theological reasons for this—then Darwinism does not refute this, but shows rather why physical pain is bound to occur.

## INTELLIGENT DESIGN

Notoriously, from the beginning many American evangelical Christians have rejected all forms of evolution. The best-known clash between such Christians and evolutionists occurred in 1925 in the state of Tennessee, when the young school teacher John Thomas Scopes was put on trial for teaching evolution. As it happened, although Scopes was found guilty, his penalty was overturned on appeal, and that was the end of so-called creationism for several decades. Yet thanks to a number of dedicated fundamentalists, people who insist on taking every verse of the Bible literally, opposition to evolutionism started to grow again, particularly after the publication in 1961 of *Genesis Flood*, a work by the biblical scholar John Whitcomb and the hydraulic engineer Henry Morris defending every verse of the Bible. This led to renewed efforts to get literal biblical teachings into publicly financed American schools, and again the matter ended in court, this time in Arkansas in 1981, where it was ruled that “creation science” is religion and as such has no place in school biology classes.

More recently, those who oppose Darwinian evolution on religious grounds have been promoting a more sophisticated form of creationism. Supporters of intelligent design argue that the organic world is just too complex and tightly functioning to have been produced by natural forces. The world, particularly at the micro level, exhibits what they call “irreducible complexity,” and hence cannot possibly have been the result of something like natural selection. In the words of Michael Behe, author of *Darwin’s Black Box*, an “irreducibly complex

biological system, if there is such a thing, would be a powerful challenge to Darwinian evolution. Since natural selection can only choose systems that are already working, then if a biological system cannot be produced gradually it would have to arise as an integrated unit, in one fell swoop, for natural selection to have anything to act on” (p. 39).

As an example of something irreducibly complex, Behe turns to the micro world of the cell and of the mechanisms found at that level. Take bacteria that use flagella, driven by a kind of rotary motor, to move around. Every part is incredibly complex, and so are the various parts in combination. For example, the flagellin (the external filament of the flagellum) is a single protein that forms a kind of paddle surface contacting the liquid during swimming. Near the surface of the cell, one finds a thickening, just as needed, so that the filament can be connected to the rotor drive. The connector is a hook protein. There is no motor in the filament, so it has to be located somewhere else. And so on. Such an intricate mechanism is much too complex to have come into being in a gradual fashion. Only a one-step process will do, and this one-step process must involve some sort of designing cause. Behe and his supporters, including the mathematician-philosopher William Dembski, are careful not to identify this designer with the Christian God, but the implication is that the designing cause is a force beyond the normal course of nature. Biology works through “the guidance of an intelligent agent” (p. 96).

Evolutionists strongly deny that there are irreducibly complex phenomena, and they strive to show that the adaptations highlighted by intelligent-design theorists could in fact have been produced by natural selection. Of course, often mechanisms as we see them today could not function if a part were removed, but this is compatible with their coming into being through blind natural law. Perhaps formerly essential but now redundant parts have been removed. Think of a stone arch. Build it without supports, and the center keystones will fall before they are secured. Build supports and then build the arch, and the completed structure will stand even after the supports are removed.

In any case, argue critics of intelligent-design theory, there are significant theological problems with the theory, which is little more than Paley’s natural theology brought up to date with some modern examples. If an intelligence intervened to produce the irreducibly complex, why does the intelligence not intervene to prevent life’s simple but devastating occurrences? Sometimes a simple change in the structure of DNA can have horrific effects on an indi-



vidual. Why are these sorts of occurrences not prevented? One might say that the intelligence is not interested in doing everything, but if this is true, then it at least seems that the intelligence pointed to by intelligent-design theory is far removed from the traditional Christian conception of God.

## CONCLUSION

There is more debate at the beginning of the twenty-first century than perhaps at any other time about the relationship between science and religion, and in particular between biology and Christianity. It is neither static nor philosophically uninteresting.

*See also* Religion and the Physical Sciences.

## Bibliography

- Behe, Michael. *Darwin's Black Box: The Biochemical Challenge to Evolution*. New York: Free Press, 1996.
- Conway Morris, Simon. *Life's Solution*. Cambridge, U.K.: Cambridge University Press, 2003.
- Darwin, Charles. *On the Origin of Species*. London: John Murray, 1859.
- Darwin, Erasmus. *The Temple of Nature*. London: J. Johnson, 1803.
- Dawkins, Richard. *The Blind Watchmaker*. New York: W. W. Norton, 1986.
- Dawkins, Richard. *A River out of Eden*. New York: Basic Books, 1995.
- Dawkins, Richard. "Universal Darwinism." In *Evolution of Molecules to Men*, edited by D. S. Bendall. Cambridge, U.K.: University of Cambridge Press, 1983.
- Gould, Stephen Jay. *Wonderful Life: The Burgess Shale and the Nature of History*. New York: W. W. Norton, 1989.
- Newman, John Henry. *The Letters and Diaries of John Henry Newman*. Vol. 25, edited by Charles Stephen Dessain and Thomas Gornall. Oxford, U.K.: Clarendon Press, 1973.
- Paley, William. *Natural Theology* (1802). Vol. 4 of his *Collected Works*. London: Rivington, 1819.
- Pannenberg, Wolfhart. *Towards a Theology of Nature*. Louisville, KY: Westminster/John Knox Press, 1993.
- Ruse, Michael. 2001. *Can a Darwinian be a Christian? The Relationship between Science and Religion*. Cambridge, U.K.: Cambridge University Press.
- Teilhard de Chardin, Pierre. 1959. *The Phenomenon of Man*. London: Collins.
- Whitcomb, John C., and Henry M. Morris. *The Genesis Flood: The Biblical Record and Its Scientific Implications*. Philadelphia: Presbyterian and Reformed Publishing, 1961.

**Michael Ruse (2005)**

## RELIGION AND THE PHYSICAL SCIENCES

This entry is concerned with philosophical questions arising from the interaction of religion and physical science. Here the focus is primarily upon Western religious monotheism, for this is the larger religious context in which modern science arose. And among the physical sciences, the focus is on astronomy and physics.

## HISTORICAL ROOTS

The relationships between physical science and monotheism have deep roots in the history of Western thought. The simple assumption that religion and science have been and remain in conflict is falsified by the historical data. Rather, more complex and interesting connections hold between religious faith and scientific understanding in at least three domains: individual scientists and scholars, social institutions, and worldviews. At the individual level, the facts are too complex for one simple view to be true all the time, or even in a majority of cases. At the institutional level, the record of religion is at best one of indifference, and at worst outright opposition to physical science. At the level of worldviews, in contrast, Western religion has helped to make modern physical science possible.

The regular pattern of astronomical events traced by ancient Babylonian astrologers and the understanding of the physical world in Greek natural philosophy and astronomy gave currency to the idea that there must be a supreme god of some sort behind the universal patterns of causes and motions in heaven and earth. As Plato argued in *The Laws*, "If the whole path and movement of heaven and all its contents are of like nature with the motion, revolution, and calculations of wisdom, and proceed after that kind, plainly we must say that it is the supremely good soul that takes forethought for the universe and guides it along that path" (bk. 10, 897c). Both Plato and Aristotle were philosophical monotheists, a view based in part on their understanding of the workings of nature.

The tradition of Greek natural philosophy continued to develop in the monotheistic traditions of Christian, Jewish, and Arabic scholarship by means of commentaries on the physical works of Aristotle. What these philosophers had in common was what we might call a *macrodesign* scientific worldview: God created the whole cosmos and sustains the principles and laws of nature that regulate physical interaction and motion. The purpose of natural philosophy (as physical science was then

called) was to investigate the primary and secondary causes sustained by the first cause. Natural philosophy did not discuss God per se as the first cause, nor did it appeal to God as an explanation for the natural phenomena of the world. God's nature was the province of theology. This division of labor aided the development of the rationality of early modern science in the European universities in the thirteenth century, the later Middle Ages, and the Renaissance.

Important to this development was the influx of the "new" Aristotelian science from Arabic sources. Combined with a Platonic-Pythagorean tradition of mathematics, this Aristotelian tradition of empirical study was assisted by voluntarism in theology and nominalism in metaphysics. This complex tradition of inquiry formed the background to the development of early modern science and made sense of a quest for empirical, mathematical laws of nature grounded in the will of the Creator. A good historical example of this combination is Jean Buridan (c. 1292–1358), a natural philosopher and one of the most honored intellectuals in Europe, who was twice elected rector of the University of Paris. In his commentary on Aristotle's *On the Heavens*, he wrote, "In natural philosophy we ought to accept actions and dependencies as if they always proceed in a natural way" (bk. 2, ques. 9; p. 423 f.). In the same question, Buridan went on to attribute the existence and design of the universe to God as first cause, but he did not appeal to God in natural philosophy.

The scientific revolution was a genuine revolution in human thought. Despite some continuity with the Middle Ages, a whole new way of seeing the world was born. The contributions of Nicolaus Copernicus, Galileo Galilei, Johannes Kepler, and Isaac Newton, for example, gave rise to a new understanding of the physical cosmos. While a macrodesign worldview did assist in the development of early modern science, there was tension at the institutional level. The Catholic Church continued to insist upon its right to judge theological truth, including the proper way to interpret the Scriptures. The Catholic astronomer Galileo Galilei (1564–1642) ran into trouble with the Congregation of the Holy Office (the Inquisition) over exactly this point. In his famous "Letter to the Grand Duchess Christina" (1615/1957), he argued as an individual scholar that the Scriptures should be interpreted in a manner consistent with the new Copernican astronomy. The Counter-Reformation authorities in Rome soon banned the work of Copernicus "until corrected," and got Galileo to agree not to publish his views except as purely hypothetical theories. When Galileo

broke this agreement by publishing his *Dialogue concerning the Two Chief World Systems*, he was suspected of heresy and forced to recant publicly. This became the most famous example of institutional religion suppressing the scientific quest for truth in the physical sciences.

For the most part the Christian churches have been unconcerned with science, focusing instead on spiritual truth and religious practices. Indeed, by creating the Western university and the hospital, the Church provided indirect support for scientific research.

## METHODOLOGY

By way of contemporary issues of philosophical interest, the rise in the latter half of the twentieth century of theology-and-science debates has stimulated a number of methodological questions concerning both religion and science. The question of how we know in both disciplines has given rise to philosophical investigation into the nature and limits of knowledge in physical science and academic theology. Thomas Kuhn's *Structure of Scientific Revolutions* (1962), a revolutionary work in the philosophy of science, made a lasting contribution to the dialogue between theology and science. Science, according to Kuhn, is based on tradition and on "paradigms" of shared values, rationalities, and perspectives that gave shape to each of the scientific disciplines. It is thus based on epistemic values and metaphysical presuppositions that it owns but cannot justify. Far from being a complete worldview, science depends upon these larger perspectives for its working assumptions. This overarching view brought science into closer contact with philosophy and religion, since it was no longer the domain of purely objective, empirical fact derived from logic and evidence alone.

Investigations of the different methods of theology and science has also raised issues in the philosophy of language. How language is used in both physical science and theology has highlighted the importance of analogy and metaphor for both disciplines (Barbour 1974). This is especially true in subjects that study phenomena beyond human experience or full comprehension, for example, God and quantum reality. Yet both theology and quantum physics wish to make truth claims about their subjects, and this can only happen if we allow metaphorical truth and analogical predication.

## MATHEMATICAL PERSPICUITY

Contemporary physical science, going back to the days of Galileo, constantly uses mathematics to model reality. Yet mathematics is a symbolic language that humans created

over centuries but never grounded in pure logic. Why should mathematics be such a powerful tool to describe physical reality? The physicist Eugene Wigner raised this question in his oft cited essay “The Unreasonable Effectiveness of Mathematics in the Natural Sciences” (1960). The structures of mathematics and the deep structure of the physical universe share a feature that makes physics possible. Especially in the area of quantum physics, the ability of mathematics to predict the outcome of difficult and complex experiments is a striking example of this aspect of the universe. For theoretical physicists, the beauty and elegance of the mathematical formulas of a theory has become a key indication of the truth of the theory. But why should this be so? Is there any a priori reason to believe that the structures of mathematics should describe and predict the nature of the cosmos so well? Religious faith, especially monotheism, provides an answer to this question. The rational mind that designed the cosmos set it on a mathematically well behaved path (macrodesign again). Whether this is *the* answer to the question is a matter of serious dispute. A possible naturalistic answer might point to the evolution of the brain. Human consciousness (including the ability to create mathematics) is the ultimate product of the very laws and principles of nature that we study—a fact that makes their harmony seem more reasonable, perhaps.

## ASTRONOMY AND COSMOLOGY

From mathematics we now turn to astronomy. Three areas of this science have especially drawn the attention of philosophers and theologians: the age and size of the universe, big-bang cosmology, and the fine tuning of certain physical constants in a way that allows for the evolution of stars, planets, and people. This discussion requires the distinction between a universe and the cosmos. Here “universe” refers to our space-time domain. A universe is a spatially related collection of objects under a set of natural laws and principles. “Cosmos” refers to all the universes that have ever been or ever will be.

Our universe is expanding, and this implies that it had a beginning, when the volume of space was zero and physical time first began. Along with this discovery, astronomers in the twentieth century discovered how vast the universe really is. We are a very small part of a gigantic system of planets, stars, galaxies, and galactic clusters whose vast reaches boggle the mind. Just our galaxy alone consists of 100 million stars, and many of them may well have planets. How can we think of the Earth or our species as special in any way? Philosophers and scientists alike have embraced a kind of Stoic defiance against a

cold, dark, empty universe in which humanity has no special place. Somehow the vastness of space and time makes humans less significant, they argue. However, this ignores the fact that the God of traditional Western religion is both eternal and omnipresent. To an infinite, unbounded deity, what difference can it make how big or old the cosmos is? Any finite being will be the same relative to the creator, namely, of limited time and size. In biblical religion, the special status of human beings comes from their capacity for a personal relationship with God, not from how big, strong, or old they are. Still, the scientific conception of our universe has forced religious scholars to rethink the interpretation of the Scriptures and their understandings of the place of humans within creation. But nothing in the size and age of the universe actually falsifies the teachings of the great world religions.

The development of the concept of an initial singularity for the entire universe is one of the fascinating stories of twentieth-century physics. Suffice it to say that reluctantly, after several decades of debate, the physics community agreed that the general structure of space-time is dynamic. While such a conception of the beginning of the universe fits very poorly with the scientific materialism common in the physics community of the twentieth century, it does fit quite well with the older macrodesign view. The problem has to do with what caused the cosmos to come into existence. Even if space and time break down at the very earliest moments of space-time, we can still point to the first instances of time (which would not have any particular metric) and ask, Where did that come from? What caused it to be? Where do the structures and laws that allow such an event to take place come from? A macrodesign worldview has an answer to these questions—not a scientific one, but a religious one. The cosmos has a creator of some kind, who must be eternal, omnipotent, and omniscient (in fact, a necessary being). Note that this answer is not physical but metaphysical. It has implications for religion as well.

Philosophers who resist this implication, such as Quentin Smith and Adolf Grünbaum (2000), are forced to suggest either that (1) the earliest prematerial phase of the first quantum field that gave rise to the big bang sprang into being from nothing at all, or that (2) we can only ask questions about things that begin to exist when there is a space-time metric to measure temporality (Grünbaum), or that (3) the cosmos was just an accidental, random event in an infinite series of random events. None of these answers is especially cogent. First, a quantum field is, after all, a kind of order. Where did this order come from? If matter is structured energy, as quantum

theory teaches, the origin of structure is the key to the question of where matter comes from. The idea that all matter sprang out of an utter nothing at all—not simply no particles, but no laws, no fields, no energies of any kind—seems rather absurd. Second, to suggest that we can only think about why things come into being when there is a temporal metric to the time in question confuses physics and metaphysics. In metaphysics, it is still perfectly natural and rational to ask where the universe and its measurable temporal passage came from (and where it came from in the first place), even if there was no physical, measurable time prior to the first event of cosmic time. Finally, to suggest that the whole cosmos is purely random seems much more like an evasion of the problem than an attempt to answer the question. To postulate an infinite number of universes (or space-time domains) only to explain the design of this one is ad hoc and violates in the most extreme way Ockham's razor, or the principle of simplicity. We should not need to be reminded that this principle is important to the rationality of both physics and metaphysics. The existence and ultimate origin of the cosmos cry out for an explanation. This final issue, however, raises the question of design, and the possibility of a "multiverse," in the fundamental structures of physical reality.

#### FINE TUNING, DESIGN ARGUMENTS, AND THE MULTIVERSE HYPOTHESIS

In addition to the cosmological argument (the existence of the universe as evidence for the existence of God), in the 1990s there appeared a new and powerful version of the design argument that relies on certain fundamental constants in nature. It seems that for any intelligent life (including human life) to ever evolve anywhere in the universe, the exact values of some fundamental physical constants must be so precisely fine-tuned and balanced that it boggles the human imagination. For this reason John Barrow and Frank Tipler have called these physical constants "anthropic."

This quality of fine-tuning for anthropic purposes is widespread. Stephen Hawking, for example, estimates that the initial temperature of the universe at  $10^{-43}$  second was fine-tuned to one part in a trillion. A tiny increase would have precluded galaxies from condensing out of the expanding matter; a tiny decrease would have resulted in the collapse of the universe. Such fine-tuning is also present in two constants in Einstein's equations for general relativity: the gravitational constant and the cosmological constant. It is also found in the fine-structure constant (which regulates electromagnetic interaction),

the proton-to-neutron mass ratio, the weak nuclear force, the strong force, and so forth. According to Barrow and Tipler, a 50 percent decrease in the strength of the strong force, to take another example, would make all elements necessary for life unstable.

The initial response to this problem was to develop a number of inflationary models of the big bang. According to such models (and there are many of them), matter in the very early universe ( $10^{-35}$  second) expanded faster than the speed of light but then slowed down, and this resulted in a nearly flat curvature of space and the isolation of our relatively homogeneous space-time within a larger cosmos. We should remember that these models are highly speculative and as yet have no empirical support (that is, they are mathematical and theoretical constructions). On the basis of some inflationary models, theoretical physicists have gone even farther and suggested that our cosmos may be a "multiverse." In such theories, which need much further investigation in both physics and metaphysics, our universe is one space-time domain in a vast cosmos that might contain a large number of other universes. No serious astronomer or physicist suggests that there are an infinite number of universes. But there could be an extremely large number of universes in the cosmos, and the number might be potentially infinite (that is, finite at any moment of time but open to an infinite future). If we assign laws and principles of nature randomly among all these universes in the cosmos, the fact that ours is so well fine-tuned for the evolution of intelligent life seems less surprising.

But is it less surprising? Stephen Barr (2003) has argued cogently that even if there are many, many universes in the cosmos, the fine-tuning needed across the whole range of principles and laws is so great that no finite number of universes would lower the "surprise" (the probability of our universe, against a background knowledge consisting only of the truths of reason). If Barr is even close to being right, then the multiverse hypothesis does very little to make our biologically friendly universe less surprising (or more probable). Perhaps some macrodesign scientific worldview is the most rational explanation of the order of the universe. Other options are possible, of course, for those uncomfortable with belief in some kind of creator. These options include the view that epistemic probabilities are purely subjective, and that the only real probabilities are physical ones, so that one simply cannot judge probabilities for the initial conditions of a universe. Another possibility is that our probability reasoning cannot apply to a whole universe: Any universe is just as improbable (and just as probable)

as the next one. We are extremely lucky that one universe in the cosmos of multiple space-time domains is capable of bearing life. Despite these options, or perhaps because of them, philosophers, scientists, and theologians continue to find the new fine-tuning arguments of great interest.

**See also** Religion and the Biological Sciences.

### **Bibliography**

- Barbour, Ian. *Myths, Models, and Paradigms*. New York: Harper, 1974.
- Barr, Stephen M. *Modern Physics and Ancient Faith*. Notre Dame, IN: University of Notre Dame Press, 2003.
- Barrow, John D., and Frank J. Tipler. *The Anthropic Cosmological Principle*. Oxford, U.K.: Clarendon Press, 1986.
- Buridan, Jean. *Ioannis Buridani Expositio et Quaestiones in Aristotelis De Caelo*, edited by Benoit Patat. Louvain, Belgium: Peeters, 1996.
- Craig, William Lane, and Quentin Smith. *Theism, Atheism, and Big Bang Cosmology*. Oxford, U.K.: Oxford University Press, 1993.
- Davies, Paul. *The Accidental Universe*. Cambridge, U.K.: Cambridge University Press, 1998.
- Denton, Michael. *Nature's Destiny*. New York: Free Press, 1998.
- Earman, John, and Jesus Mosterin. "A Critical Look at Inflationary Cosmology." *Philosophy of Science* 66 (1999): 1–49.
- Galilei, Galileo. *Dialogue concerning the Two Chief World Systems, Ptolemaic and Copernican*. Translated by Stillman Drake. Berkeley: University of California Press, 1953.
- Galilei, Galileo. "Letter to the Grand Duchess Christina" (1615). In *Discoveries and Opinions of Galileo*. Translated by Stillman Drake. Garden City, NY: Doubleday, 1957.
- Grünbaum, Adolf. "A New Critique of Theological Interpretations of Physical Cosmology." *British Journal for the Philosophy of Science* 51 (2000): 1–43.
- Guth, Alan. *The Inflationary Universe*. New York: Helix Books, 1997.
- Hawking, Stephen. *A Brief History of Time*. New York: Bantam, 1988.
- Kaiser, Christopher. *Creational Theology and the History of Physical Science*. Leiden, Netherlands: Brill, 1997.
- Kuhn, Thomas. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1962.
- Leslie, John, ed. *Physical Cosmology and Philosophy*. New York: Macmillan, 1989.
- Lindberg, David C., and Ronald L. Numbers, eds. *God and Nature: Historical Essays on the Encounter between Christianity and Science*. Berkeley: University of California Press, 1986.
- Manson, Neil A., ed. *God and Design*. London: Routledge, 2003.
- Padgett, Alan G. "Science and Theology." In *The Encyclopedia of Christianity*, edited by Erwin Fahlbusch et al., 4: 192–198. Grand Rapids, MI: Eerdmans, 2006.
- Padgett, Alan G. *Science and the Study of God*. Grand Rapids, MI: Eerdmans, 2003.

- Polkinghorne, John. *The Faith of a Physicist*. Princeton, NJ: Princeton University Press, 1994.
- Shea, William, and Mariano Artigas. *Galileo in Rome*. Oxford, U.K.: Oxford University Press, 2003.
- Swinburne, R. G. *The Existence of God*. 2nd ed. Oxford, U.K.: Clarendon Press, 2004.
- Wigner, Eugene P. *Symmetries and Reflections*. Bloomington: Indiana University Press, 1967.
- Wigner, Eugene. "The Unreasonable Effectiveness of Mathematics in the Natural Sciences." *Communications in Pure and Applied Mathematics* 13 (1) (1960).

**Alan G. Padgett (2005)**

## RELIGIOUS EXPERIENCE

Most of the philosophical work on "religious experience" that has appeared since 1960 has been devoted to its phenomenology and epistemic status. Two widely shared assumptions help account for this—that religious beliefs and practices are rooted in religious feelings and that whatever justification they have largely derives from them.

The majority of the discussions of the nature of religious experience are a reaction to Walter Stace, who believed that mysticism appears in two forms. Extrovertive mysticism is an experience of nature's unity and of one's identity with it. Introvertive mysticism is an experience of undifferentiated unity that is devoid of concepts and images; it appears to be identical with what others have called "pure consciousness"—a state in which one is conscious but conscious of nothing.

R. C. Zaehner argued that Stace's typology ignores love mysticism in India and the West. There are two types of introvertive mysticism—monistic (pure consciousness) and theistic. The latter is a form of mutual love that unites God and the mystic in an experience without images and with very little, if any, conceptual content. The most effective defense of a position of this sort is Nelson Pike's. Pike argues that the principal forms of mystical prayer in Christianity (quiet, rapture, and full union) are phenomenologically theistic. He defends his analysis against William Forge, who denies that the identification of the experience's object with God can be part of its phenomenological content.

Phenomenological analyses of religious consciousness presuppose that we can distinguish descriptions of religious experience from interpretations. Ninian Smart proposed two tests for distinguishing descriptions—that the accounts be autobiographical and that they be relatively free from doctrinal concepts. The question of crite-

ria remains vexed, however (see Wainwright, 1981, chap. 1).

Others have argued that, because religious experience is significantly constituted by the concepts, beliefs, expectations, and attitudes that the mystic brings to it, attempts to distinguish interpretation from description are misguided. For example, an influential article by Steven Katz contends that a mystic's experiences are largely shaped by his or her tradition. This has two consequences. First, there are no "pure" or "unmediated" mystical experiences and, second, there are as many types of mystical experiences as there are traditions.

Katz's "constructivism" has been attacked by Robert Forman and Anthony Perovitch among others. Since pure consciousness is devoid of content, it is difficult to see how it could be constituted by contents that the mystic brings to it. To argue that it must be mediated because all experience is mediated begs the question; on the face of it, pure consciousness is a counterexample to the thesis in question. Forman also argues that constructivism cannot adequately account for novelty—the fact that the mystic's experiences are often unlike what he or she expected.

Defenses of religious experience's cognitive validity have taken several forms. William Wainwright argues that mystical experiences are presumptively valid because they are significantly similar to sense experiences. Both experiences have what George Berkeley called "outness"—the subject has the impression of being immediately presented with something transcending his or her own consciousness. Corrigible and independently checkable claims about objective reality are spontaneously made on the basis of both types of experience. There are tests in each case both for determining the reality of the experience's apparent object and for determining the genuineness of apparent perceptions of it. The nature of the tests, however, is determined by the nature of the experiences' alleged objects. Since the apparent objects of religious experience and ordinary perceptual experience differ, so too will the tests for veridical experiences of those objects.

Richard Swinburne's defense of religious experience's cognitive validity is based on the principle of credulity, which roughly states that apparent cognitions are innocent until proven guilty. This is a basic principle of rationality; without it we would be unable to justify our reliance on memory, sense perception, and rational intuition. The principle implies that there is an initial presumption in favor of how things seem to us, although this presumption can be overridden. What is true of apparent cognitions in general is true of religious experiences. They too should be accepted in the absence of good

reasons for thinking them deceptive. Swinburne argues that there are none.

The most sustained defense of religious experience's epistemic credentials is William Alston's. Whereas Wainwright and Swinburne concentrate on perceptual (or perception-like) experiences, Alston focuses on perceptual practices. Doxastic (belief-forming) practices are basic when they provide our primary access to their subject matter. The reliability of a basic doxastic practice like memory cannot be established without circularity; any attempt to justify it relies on its own outputs. Alston argues that sense-perceptual practice and "Christian mystical practice" are epistemically on a par. Since both doxastic practices are basic, neither's reliability can be established without circularity. Both practices are socially established, internally consistent, and consistent with the outputs of other well-established practices. They are also self-supporting in the sense that they have the outputs we would expect them to have if they were reliable (successful predictions in the first case, for example, and moral and spiritual improvement in the second). Alston concludes that it is unreasonable to engage in sense-perceptual practice while rejecting the rationality of engaging in Christian mystical practice. The rationality at issue, however, is not epistemic. Neither practice can be shown to be epistemically rational, since it is impossible to establish their reliability without circularity. Alston intends to show only that it is practically or pragmatically rational to engage in them, although it should be noted that engaging in them involves accepting their outputs as true and therefore believing that they are reliable. Alston concedes that the existence of competing mystical practices weakens his case but denies that it destroys it. Critiques of Alston's work have tended to focus on this point (see, for example, Hasker, 1986).

The most significant attacks on religious experience's cognitive validity to have appeared since 1960 are Wayne Proudfoot's and Richard Gale's. Proudfoot argues that an experience's noetic quality should be identified with its embedded causal judgment (that the experience is caused by a tree, for example, or by God) and this judgment's affective resonance. The incorporated causal judgment has no intrinsic authority; it is merely one hypothesis among others and should be accepted only if it provides a better overall explanation of the experience than its competitors'. While the causal hypotheses embedded in religious experiences could be correct, they are in fact suspect; they appear to be artifacts of the subject's religious or cultural tradition and not products of nonnatural causes.

Proudfoot's identification of an experience's noetic quality with an incorporated causal judgment and its affective resonance is more plausible in some cases than others. Given my background knowledge, I believe that a certain sort of pain in one's tooth is caused by cavities. Believing this, and having a pain of that sort, I spontaneously form the belief that my pain is caused by a cavity. While my pain is not noetic, the experience as a whole is, since it incorporates a causal judgment. But the experience lacks "outness." It thus differs from sense perception, which (because of this quality) seems to have an intrinsic authority that noetic experiences like my toothache lack. Religious experiences are also diverse. Some, like my toothache, involve spontaneous causal attributions and nothing more. Others, however, are perception-like and have the same claim to intrinsic authority that sense perceptions do.

Richard Gale, on the other hand, argues that religious experience lacks the authority of sense experience. The only way of establishing religious experience's cognitiveness is by showing that the tests for it are similar to those for sense experience. Arguments for religious experience's cognitive validity fail because the dissimilarities are too great. Alston and Wainwright contend that these dissimilarities can be explained by differences in the experiences' apparent objects. Gale objects that explaining the disanalogies does not explain them away and that there is a "tension" or "inconsistency" in claiming that the tests are similar (as they must be if the defense of religious experience's cognitiveness is to be successful) and yet different in nature. The first point is dubious. Only relevant disanalogies count. The point of Wainwright's and Alston's explanations is to show that the disanalogies are not relevant—that is, that the features that tests for sense experiences have and tests for religious experiences lack are not ones we would expect the latter to have if religious experiences were veridical perceptions of their apparent objects.

Gale's most original (and controversial) contribution is his contention that veridical experiences of God are conceptually impossible. The argument is roughly this: Talk of veridical experiences is in place only where it makes sense to speak of their objects as existing "when not actually perceived" and as being "the common object of different" experiences of that type. Sense experiences exhibit this feature because their objects are "housed in a space and time that includes both the object and the perceiver." Religious experiences do not exhibit this feature because there are no "analogous dimensions to space and time" that house both God and the perceiver. Gale

attempts to establish this by refuting P. F. Strawson's claim that a "no space world ... of objective sounds" is conceptually possible. We could neither reidentify sounds in such a world nor distinguish between numerically distinct but qualitatively identical ones. It would make no sense, therefore, to speak of sounds as the common objects of distinct auditory experiences or as existing when unperceived. Talk of veridical experiences of objective sounds would thus be out of place. A fortiori, talk of veridical experiences would be out of place in a nonspatial and nontemporal world. Therefore, since no common space (and, on some accounts, no common time) houses God and the mystic, talk of veridical perceptions of God is inappropriate.

A few general observations about discussions of religious experience since 1960 are in order. First, most defenses of religious experience's cognitive validity have been offered by theists. Stace is one of the few who has attempted to establish the veridicality of pure consciousness and other nontheistic experiences that lack intentional structure. Second, philosophical discussions of religious experiences tend to abstract them from the way of life in which they occur and thereby impoverish our understanding of them. Whether this penchant for abstraction adversely affects the discussion of phenomenological and epistemological issues, however, is more doubtful. Finally, a philosopher's assessment of the cognitive value of religious experience is affected by his or her metaphysical predilections. For example, those who assign a low antecedent probability to theism will demand stronger arguments for theistic experiences' cognitive validity than those who do not. One's assessment of religious experience cannot be separated from one's general assessment of the relevant religious hypotheses.

**See also** Alston, William P.; Berkeley, George; Constructivism and Conventionalism; Constructivism, Moral; Intuition; Memory; Mysticism, History of; Mysticism, Nature and Assessment of; Perception; Philosophy of Religion; Rationality; Stace, Walter Terence; Strawson, Peter Frederick.

### **Bibliography**

- Alston, W. P. *Perceiving God*. Ithaca, NY: Cornell University Press, 1991.
- Davis, C. F. *The Evidential Force of Religious Experience*. Oxford: Clarendon Press, 1989.
- Forgie, W. J. "Pike's Mystic Union and the Possibility of Theistic Experience." *Religious Studies* 30 (1994): 231–242.
- Forgie, W. J. "Theistic Experience and the Doctrine of Unanimity." *International Journal for Philosophy of Religion* 15 (1984): 13–30.

- Forman, R. K. C. "Introduction: Mysticism, Constructivism, and Forgetting." In *The Problem of Pure Consciousness*, edited by R. K. C. Forman. New York: Oxford University Press, 1990.
- Gale, R. *On the Nature and Existence of God*. Cambridge, U.K.: Cambridge University Press, 1991. Chap. 8.
- Gutting, G. *Religious Belief and Religious Skepticism*. Notre Dame, IN: University of Notre Dame Press, 1982. Chap. 5.
- Hasker, W. "On Justifying the Christian Practice." *New Scholasticism* 60 (1986).
- Katz, S. T. "Language, Epistemology, and Mysticism." In *Mysticism and Philosophical Analysis*, edited by S. T. Katz. London: Sheldon Press, 1978.
- Perovitch, A. N., Jr. "Does the Philosophy of Mysticism Rest on a Mistake?" In *The Problem of Pure Consciousness*, edited by R. K. C. Forman. New York: Oxford University Press, 1990.
- Pike, N. *Mystic Union: An Essay in the Phenomenology of Mysticism*. Ithaca, NY: Cornell University Press, 1992.
- Proudfoot, W. *Religious Experience*. Berkeley: University of California Press, 1985.
- Smart, N. "Interpretation and Mystical Experience." *Religious Studies* 1 (1965): 75–87.
- Stace, W. T. *Mysticism and Philosophy*. Philadelphia: Lippincott, 1960.
- Swinburne, R. *The Existence of God*. Oxford: Clarendon Press, 1979. Chap. 13.
- Wainwright, W. J. *Mysticism: A Study of Its Nature, Cognitive Value, and Moral Implications*. Madison: University of Wisconsin Press, 1981.
- Yandell, K. E. *The Epistemology of Religious Experience*. Cambridge, U.K.: Cambridge University Press, 1993.
- Zaehner, R. C. *Concordant Discord*. Oxford: Clarendon Press, 1970.

William J. Wainwright (1996)

## RELIGIOUS EXPERIENCE, ARGUMENT FOR THE EXISTENCE OF GOD

Arguments from Religious Experience show remarkable diversity, (a) in the sorts of experience taken as data for the argument, (b) in the structure of the inference itself, and (c) in the alleged conclusion, whether to a vague Presence, an Infinite Being, or the God of traditional Christianity.

The following exemplify some versions of the argument:

"At very different times and places great numbers of men have claimed to experience God; it would be unreasonable to suppose that they must all have been deluded."

"The real argument to God is the individual believer's sense of God's presence, the awareness of God's

will in tension and conflict with his own will, the peace that follows the acceptance of God's command."

"Experiences of meeting God are self-authenticating: They involve no precarious chain of inference, no sifting of rival hypotheses. They make unbelief logically absurd."

"In itself, religious experience is neither theistic nor pantheistic, Christian nor Buddhist. All these distinctions are *interpretations* of the experience. By itself, religious experience testifies to something far less definite but still infinitely valuable—the insufficiency of all materialisms and naturalisms."

If we compare any of these arguments with the Ontological, Cosmological, and Teleological arguments, important differences in their logic and history can readily be shown. Arguments from Religious Experience are clearly not a priori, like the Ontological Argument, and whereas the Cosmological and Teleological arguments work from premises that affirm highly general facts about the world (that it exists, that it is purposefully ordered), Arguments from Religious Experience rely on far more particularized and elusive premises than these. Not all men have (or are aware of having) distinctively religious experiences, and to those that do have them religious experiences are apt to be short-lived, fugitive sets of events that are not publicly observable.

Despite this slipperiness, the Argument from Religious Experience has attracted some theologians who have been skeptical about the more rationalistic "proofs." In the course of the eighteenth century these proofs received formidable criticism from Immanuel Kant and David Hume. The Ontological Argument was shown to be radically confused over the logic of "existence," and (in Kant's account) the Cosmological and Teleological arguments themselves presuppose the Ontological. Even more important, Kant and Hume together produced a general weakening of confidence that any survey of the observable cosmos (including "the starry heavens above") could yield premises powerful enough to argue to an infinite, unconditioned, all-good deity. Kant turned to "inner" experience, to our awareness of the moral law, and argued that the moral life is intelligible only if we postulate God and immortality.

Although a number of writers followed Kant in arguing from inner moral experience, many others, while accepting the shift from outer to inner, based their inference on a distinctive class of religious experiences. If we describe this shift, in general terms, as a move from objective to subjective, from surveying the world at large for evidences of God to focusing attention on the personal



and existential, it clearly was a shift of the greatest moment and one that still helps to determine our contemporary climate of theological thought. We human beings are not stars or electrons—the argument goes—and we cannot experience or guess the role of star or electron in the divine economy. But we are persons, and we are directly aware (or some of us are) of a meeting of person with Person in religious experience.

Thinking back, however, to the post-Humean, post-Kantian period, the centering upon inner experience can be seen as one aspect of the romantic movement's protest against the Enlightenment, the new concern for subjectivity, the life of the emotions and intuitions of the individual. The most important and most seminal single figure here is Friedrich Schleiermacher (1768–1834), with his bold insistence on the primacy of religious feeling—particularly the feeling of utter creaturely dependence—and his distaste for religious doctrines or arguments entertained in a purely intellectual manner, as mere ideas, lacking the life and authority of experience.

## OBJECTIONS TO THE ARGUMENT

Prima facie it seemed a reasonable and empirically sound enterprise to establish arguments for God upon claims to have actually experienced him, to have “seen” him, “met” him, encountered him in a personal relation. But there are in fact several directions from which it can be challenged.

Orthodox and neoorthodox theologians tend to object that the content of religious experience is too indeterminate to yield clear knowledge of the God of Christianity. The case for Christianity must not be allowed to rest on the deceptive and elusive emotions of religious people. It rests on the revealed Word of God, on the Person of Jesus Christ as disclosed in the Scriptures, not as constructed out of the assorted emotions of the devout. The working of the Holy Spirit cannot be correlated with the experiencing of peculiar feelings, even uplifting ones.

A second familiar objection is that although we certainly do have religious experience, we cannot employ it as the premise of an argument to God. The relationship between man and God—an I-Thou (in Martin Buber's phrase), personal relationship—is maintained by faith alone. The conception of superseding faith through a proof of God's existence forgets the irreducibly personal nature of encounter between man and God.

The objector may be making a religious claim, that it is religiously improper to attempt to replace faith by rational argument, or his point may be a logical one, that

God—being “pure” person, having nothing bodily or thinglike in his nature—cannot be shown to exist in the way things can be shown to exist.

Suppose, again, we take the Argument from Religious Experience as an explanatory hypothesis; then a skeptical critic may deny that the existence of God is the likeliest, or simplest, or most intelligible, explanation of the experiences. We cannot be intellectually compelled to posit God if more economical and naturalistic explanations can be found—psychoanalytic accounts, it might be, or accounts in terms of individual suggestibility or the influence of religious expectations or tradition.

Last, a critic may concentrate on the conceptual difficulties in the idea of God, for if the argument as a whole is to be sound, its conclusion (“therefore God exists”) must be intelligible and free of inner contradictions. This objection may bewilder and disappoint the arguer-from-experience. To him one of the chief apparent advantages in the argument is that its direct appeal to experience bypasses logical or metaphysical complexities. But some element of interpretation, and therefore some application of concepts, must take place when an experience is taken to be an encounter with God. Wherever concepts are handled, they can also be mishandled. Inner contradictions in the claim to experience God could invalidate the interpretation of the experience.

## NATURE OF RELIGIOUS EXPERIENCES

What, more exactly, are religious experiences? Descriptions of religious experiences can be heavily loaded with doctrinal, even sectarian, interpretation or can be almost entirely free of it. Their impact may fix one's attitudes and evaluations for a lifetime or for only a brief period. They may not only be benign and optimistic, as we have so far assumed, but can also—with no less intensity—be pessimistic and grim. They may involve conversions to a religious orthodoxy or conversions away from one. Consider the following experiences, neither of which is more than minimally interpreted, and both of which are certainly in an important sense religious. The first is from Lev Tolstoy's *War and Peace*, at the point where Prince Andrew has been wounded in the Battle of Austerlitz.

He opened his eyes, hoping to see how the struggle of the Frenchmen with the gunners ended.... But he saw nothing. Above him there was now nothing but the sky—the lofty sky, not clear yet still immeasurably lofty, with grey clouds gliding slowly across it. “How quiet, peaceful and solemn, not at all as I ran,” thought Prince Andrew—“not as we ran shouting and fight-

ing.... How was it I did not see that lofty sky before? And how happy I am to have found it at last! Yes! All is vanity, all falsehood except that infinite sky. There is nothing, nothing but that. But even it does not exist, there is nothing but quiet and peace. Thank God!”

The second is from Leonard Woolf’s autobiographical work, *Sowing* (1960). At the age of eight, the author was sitting in a garden enjoying the fresh air after a train journey. He watched two newts basking in the sun.

I forgot everything, including time, as I sat there with those strange, beautiful creatures, surrounded by blue sky, sunshine, and sparkling sea. I do not know how long I had sat there, when, all at once, I felt afraid. I looked up and saw that an enormous black thunder cloud had crept up and now covered more than half of the sky. It was just blotting out the sun, and, as it did so, the newts scuttled back into their hole.... I felt something more powerful than fear, once more that sense of profound, passive, cosmic despair, the melancholy of a human being eager for happiness and beauty, powerless in face of a hostile universe.

Turning to theistic types of experience, we can start from the very basic experience of wonderment, notably wonderment at there being any world at all. This may pass into the sense that the world owes its existence to, and is maintained in existence by, something “beyond,” “outside” the world itself, a Being whose nature is utterly remote from the world, yet whose activity and energy are perceptible within the world, as a disturbing, awesome, and thrilling presence. Rudolf Otto’s concept of the “numinous” gathers together these ingredients of mystery, dread, and fascination and emphasizes very properly the qualitative distinctiveness and elusiveness of such experience (*The Idea of the Holy*, passim). No set of categories can neatly contain it: The person who has never known it can barely understand the claims of the person who has.

Religious experiences can be generated by perceptions of individual objects (a grain of sand, a bird), by a train of events, by actions—for instance, the memorable account of Jesus setting his face to go to Jerusalem to his Passion. Even a passage of philosophical reasoning may do this, as when someone contemplates the incompleteness of all explanation, the intellectual opacity of space and time, and feels compelled—with a sense of mystery—to posit a divine completeness and unity.

Closer to the province of morality are experiences of divine discontent, interpreted as intimations of God’s existence and call to moral endeavor, the conviction of sin correlative to a sense of God’s own holiness, the sense of divine aid in the rectifying of one’s moral life, and, in Christian evangelical terms, a sense that one has been redeemed or saved by God’s action on man’s behalf.

The overall impression is of the immense diversity of religious experiences. They are indeed linked by complex webs of “family resemblances” (to use Ludwig Wittgenstein’s phrase)—resemblances of attitude, emotional tone, alleged content—but if we ask what all of them have in common, the answer must be meager in content: perhaps only a sense of momentous disclosure, the sense that the world is being apprehended and responded to according to its true colors. What is actually being observed or contemplated can never (logically) be the whole universe, yet the quality of religious experience is such that it does seem to imply something about the whole.

#### EPISTEMOLOGICAL STATUS OF RELIGIOUS EXPERIENCES

Our sampling of religious experiences may help to deliver us from the dangers of oversimplification, but it cannot by itself determine whether arguments to God based upon them are valid. Clearly, not all the experiences we have mentioned could yield data with which a theistic argument could start. Some, such as that of Woolf quoted earlier, are decidedly *antitheistic*. But there is a further set of differences among them that must be noted at this stage, differences of an epistemological kind.

When someone speaks of his religious experience, he may be using the word *experience* as it appears in such phrases as “business experience,” “driving or teaching, etc., experience.” He has found the religious pattern of life viable; he has interpreted a multiplicity of events in its categories, and these categories have proved durable. There is the suggestion that the person with religious experience in this sense has been confirming his faith by living it out over a substantial stretch of his life—furnishing data for a pragmatic proof of God’s existence.

In other cases the experiences are of much shorter duration, often judgments or quasi perceptions accompanied by certain religious emotions, alleged cognitive acts or intuitions in which the necessity of God’s existence is “seen” and an awesome emotive response is elicited simultaneously. Again, the language used may be nearer to that of perceiving—*seeing God* (not just seeing *that* God exists). There is a claim to knowledge of God by “acquaintance,” rather than “description.”

Some cases resemble the dawning of an aspect or interpretation, as when we recognize a person in a poor light or make out a pattern in what looked like a maze of lines. It can be like a sudden reading of the expression on a face, the face, as it were, of the universe, or like a realization of meaning, as when one sees the point of a poem with which one has long been verbally, but only verbally, familiar. In the light of this disclosure, a new orientation and purposeful organization of life may take place. Energies hitherto dissipated or in mutual conflict are rallied and integrated.

Feelings or emotions may predominate in religious experience, but even so, perception and judgment are almost always involved as well. Feelings are often “feelings *that ...*,” surmises, and in that sense feelings involve judgment, have an essential component of belief. Part of what it is to have an emotion is to see and appraise one’s situation in a particular way. (“I feel remorse for doing *x*,” for example, presupposes “I did *x* freely” and “*x* was morally wrong.”) It is only with twinges, frissons, aches, and such like that no appraisal of the situation need (logically) be made; these, in any case, could furnish only very weak premises for a theistic argument. Their occurrence can be due to a great variety of causes immanent in one’s own organism and one’s environment, and they can hardly, without supplementation, force one to posit a transcendent cause.

Obviously the structure (and maybe the validity) of an Argument to God from Religious Experience will vary enormously according to what epistemological type of experience is taken as the starting point, and in the literature this is often hard to discern.

## VERIFIABILITY OF RELIGIOUS EXPERIENCE

If someone claims to have discovered, perceived, become aware of an ordinary sort of object, we usually know what to do about checking his claim. If we are told that there is a frog in the garden pond, we know what it will be like to confirm this or to find it untrue. We know how to investigate whether it was Smith we saw in the dim light, whether we did hit the right answer to a sum or cried “Eureka” too soon. But when someone claims to have direct awareness of God, to encounter, see, or intuit the divine, we are not able to suggest a test performance of an even remotely analogous kind. The more developed and theologically sophisticated the conception of deity is, the more it eludes and resists any such check.

This being so, some critics have pointed out a disturbing resemblance between claims to experience God

and a certain other range of statements that are not publicly testable—namely, psychological statements such as “I seem to hear a buzzing noise,” or “I seem to see a patch of purple.” If statements like these cannot be refuted, it is only because they make no assertions about what exists, beyond the experiences of the speaker at the moment he speaks. But the person who says he has direct and certain experience of God wishes to claim irrefutability and to affirm at the same time something momentous about what exists. Can this be done? Or would it take a far more elaborate and many-stranded apologetic to give effective backing to these claims—especially the claims to objectivity?

One might try to obtain this support by compiling records of numerous experiences of the same general kind and treating them as cumulative evidence for the truth of claims to experience God. Without doubt there is an impressive mass of such records within the Judeo-Christian tradition. Other religious traditions, however, can also produce their own very different records—of the various well-ordered phases in the quest for nirvana or for mystical union with a pantheistic object of worship.

Are these differences, however, real incompatibilities; do they correspond to genuinely different sorts of religious experience? Or are the experiences basically the same, though differently interpreted? On this it is extremely hard to give any confident answer. Part of the difficulty is that most of the developed religions contain several strands in their conceptions of the divine. Christianity, for instance, seeks to unite numinous and mystical views of God: God is “remote” and “other,” yet also mystically “near.” What can be said again is that any common elements must be very indeterminate in content and able to bear great variety of interpretation—to be taken, among other things, as the disclosure of a state or spiritual goal (nirvāṇa) or of a personal or suprapersonal God. We have seen how an experience may have a minimal—quite undoctinal—interpretation and yet be religious in a broad sense. But from such an experience alone one can hardly infer anything so definite as the God of theism. Unfortunately, the interpretations that supplement the experience are conceptually intricate and involve all the uncertainty and fallibility of philosophical and theological speculation. In this region we are far removed from the ideals of immediacy, directness, and self-evidence.

Yet a critic who claimed that the Argument from Religious Experience was thereby refuted would be missing the mark. The theist could insist that a much too crude notion of “interpretation” has so far been used, one

that suggests, falsely, that there is a merely external and almost arbitrary relation between having and interpreting an experience.

The full impressiveness of the theistic case appears only when we survey the historical development of religious experience in the direction of Christian monotheism. As the idea of deity evolves, from finite and local numen to infinite and omnipresent Lord, from the god of a tribe or nation to the Ruler of all nature, from the deity concealed in holy tent or temple to the one God beyond all phenomena whatever, religious experience is itself simultaneously transformed. It is transformed not haphazardly but so as to produce a crescendo of numinous intensity, a constant refining away of merely superstitious and idolatrous awe at objects unworthy of worship, and the arrival of a distinctive, lofty note of adoration. Experience and interpretation here advance in indissoluble unity. It is argued that this historical development provides material for a more adequate argument to God—one in which the risks of fantasy and subjectivism are much reduced.

Impressive this is, and it may well be the truth of the matter. We must notice, however, that we are now looking at a much more complex piece of argument than the claims of individuals to have direct experience of God. New logical problems appear at several points. Can we be confident, for instance, that an intensification of numinous experience is necessarily a sign that we have a more adequate disclosure of God and not simply that we have constructed a more adequate and awesome idea of God? (This is the question that also calls in doubt any purely pragmatic philosophy of religion.) Again, sometimes an artist, or a school of artists, succeeds in progressively clarifying and intensifying an original vision or the expression of some distinctive emotion. But success in this (“now he has brought the theme to full explicitness,” for example) is not necessarily correlated with a progress in discovery about the world. Can we be sure that the development of numinous awareness is different in this vital respect?

The person with theistic religious experience is assured that it is different. But the sense of assurance, the “Aha!” experience, the penny dropping, the light dawning—these are very unreliable guides to truth, validity, or value. Not the most tempestuous sense of poetic inspiration can guarantee that a good poem is being brought to birth, nor can any of these conviction-experiences by itself authenticate its related judgments. It is enough to recall how often incompatible judgments are made with equal assurance on each side. Yet it is not easy to formu-

late a version of the Argument from Religious Experience that does not rely crucially on a sense of conviction. Even when appeal is made to the pattern of development toward theism, and thus to a far wider range of phenomena than in any argument from the experience of an individual, still the issue of objectivity—that we are coming to know God, not simply an idea of God—seems to hang upon the fallible, illusion-prone assurance of the subjects. On the other hand, to point this out is to draw attention only to the risk, not to the certainty, of being wrong. A religious person may realize, and be prepared to accept, this measure of risk.

Could we escape the uncertainty, by claiming that genuine experience of God is necessarily followed by a godly life, whereas illusory experiences betray themselves by the absence of any practical fruits? Hardly; there might well be a positive correlation between genuine experiences and godliness, but in fact they are not necessarily related. Lapses, moral failures, are always open to human beings, and one cannot rule out by definition the possibility of a man’s being both morally remarkable and atheistic.

But, one might argue, is the situation vis-à-vis God any worse in principle than the situation vis-à-vis material objects, such as tables and chairs? Our traffic with these consists in having actual experiences (visual, tactual, etc.) and ordered expectations of future and possible experiences. Where our experience has this sort of structure and can thus be the subject of intelligible discourse, we confer on it the status of objectivity without more ado. But theistic experience certainly occurs, and it too has its structure of expectations.

If we can bring out the difference between these cases (and the peculiar difficulty of the religious case), we shall be showing more clearly than hitherto that the Argument from Religious Experience is most intimately involved in problems of logic and meaning—problems that at first seem alien to its empirical appearance. With a material object (say, a cube) there are quite intricate but intelligible ways in which we come to see it as a single object out-in-the-world. It is given unity most obviously by possession of perceptible limits and boundaries and by the manner in which its several surfaces can be seen and felt to connect with one another. Moreover, we have mastered the laws of perspective and so can anticipate and understand variations in our perceptions of the object, owing to our own variable positions as observers. Such variations do not, therefore, impugn the assertion that the object exists in the world external to us.

With God, who is not a finite material object, there can be no inspecting of boundaries or surfaces. And if part of what we mean by “God” is “an infinitely and eternally loving Being,” no conceivable experience or finite set of experiences could by itself entitle us to claim that we had experienced such a being. We might well report experiencing “a sense of immense benevolence toward us,” “a sense of complete safety and well-being,” but from their intensity alone one could not rigorously conclude, “Therefore I am in touch with an infinitely and eternally loving God.” From the intensity of a human love one cannot infer, “This love will endure,” and without bringing in a supplementary doctrine of God’s attributes (not derived from experience) one could no more legitimately do so in the religious case.

Material objects, of course, are sometimes observed in unfavorable perceptual conditions—at a great distance, half-concealed, and so on. Imagination must “fill in” the perceptual gaps as best it can, until conditions improve. Analogous thought models are indeed employed in theological discourse, but they are peculiarly difficult to assess. The Christian theologian is normally most ready to admit that we can neither perceive nor imagine how the various attributes of God unite in a single being (if he is to be called “a being” at all). A fair measure of agnosticism here is compatible with full Christian belief. But it may not be compatible with a reliance upon an Argument to God from Religious Experience, if this is one’s chief apologetic instrument. Unless the principles that confer unity and objectivity are able to be collected from the experiences themselves (which seems not to be possible), we have to look elsewhere for them, and the argument is in this respect shown to be inadequate. But it is not, on that account, proved useless, for if it cannot demonstrate the existence of God unaided, it might still function as a necessary auxiliary of other arguments—for example, the Cosmological Argument.

One might be forced to a deeper agnosticism than that to which we have just alluded—deeper in that it dares to affirm scarcely anything at all about the focus (or focuses) of religious experience, whether personal or impersonal. Yet with a minimal ontological commitment it might still set great value on certain religious experiences and seek after them. The attempt to work out a coherent and systematic theological interpretation would be quite abandoned.

This would save something, but assessing just how much to expect from a religious agnosticism like this would be a difficult task. The bigger the area of agnosticism, the smaller the area of legitimate religious expecta-

tions, such as that of ultimately seeing God “face to face” or of being received by him into glory. As we have more than once observed, the relation between experience and what the experience is taken to be is a most intimate one; the experiences of a Christian and those of a religious agnostic could both be valuable but could not be identical.

## PSYCHOLOGICAL EXPLANATIONS

Is it not more enlightened, however, to deny that these experiences really disclose anything about the world? Psychoanalytic research has, after all, revealed many situations in which interior mental events are projected upon the world and are furnished with all the assurance of objectivity, the full sense of “givenness.” One does not have to accept the entire Freudian account of religion to see plausibility in its central claim that early parent-child relations of “creaturely” dependence and reverence, with their tensions between love and fear, can yield the unconscious material from which experiences of God-man relations are fashioned. To accept this claim is not necessarily to reduce all religion to neurosis or worse. For it is absurd to class together the person who attains a stable religious solution to his conflicts and the person who retreats to genuine neurosis, developing, say, obsessions, compulsions, or delusions of persecution. Sigmund Freud certainly went further in his naturalistic explanation of religious experience, being prepared to reduce God to an illusory parent substitute. It may be possible, however, to invert the Freudian account of religious experience and, instead of seeing God as a father substitute, to see human fathers as God substitutes and the human experience of love as training for loving God. The close psychological relation between love of man and love of God would thus have its skeptical sting removed. It may be argued, again, that naturalistic and Christian explanations are compatible: God may elicit from us an effective response to his existence without making use of anything but our natural human equipment of senses, desires, emotions. Even mechanisms of projection can be involved and the projected image of deity be yet a trustworthy symbol of a God who does in fact exist. It is clear from all this that depth psychology does not provide a self-sufficient, decisive refutation of theism.

Nonetheless, depth psychology troubles and disturbs the Arguments from Religious Experience, and so do the very attempts to reconcile it with Christian belief. These virtually admit that the religious experiences might occur much as they actually do occur—without there being a God—in other words, that naturalistic explanations are

possible. There seems no way, at the experiential level, of settling the really urgent questions, most of all the following: Do we have in theistic experience mere projection? Or do we have a projection matched by an objectively existing God?

**See also** Agnosticism; Buber, Martin; Cosmological Argument for the Existence of God; Enlightenment; Hume, David; Kant, Immanuel; Mysticism, Nature and Assessment of; Ontological Argument for the Existence of God; Otto, Rudolf; Popular Arguments for the Existence of God; Religious Experience; Schleiermacher, Friedrich Daniel Ernst; Teleological Argument for the Existence of God.

## Bibliography

### WORKS ON THE PSYCHOLOGY OF RELIGION

The following works have important implications for the Argument from Religious Experience:

- Freud, Sigmund. *The Future of an Illusion* (1927). Translated by W. D. Robson-Scott. London: Hogarth Press, 1928; paperback ed., Garden City, NY, 1957.
- Freud, Sigmund. "Obsessive Acts and Religious Practices" (1907). In *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, translated by James Strachey, ed., in collaboration with Anna Freud. 24 vols. New York: Macmillan, 1953–1963.
- Freud, Sigmund. *Totem and Taboo* (1913). Translated by A. A. Brill. New York: New Republic, 1927. Also translated by James Strachey. New York: Norton, 1952.
- James, William. *Varieties of Religious Experience*. London: Longman, 1902.
- Jung, C. G. *Psychology and Religion*. New Haven, CT: Yale University Press, 1938; paperback ed., 1960.
- Philp, H. L. *Freud and Religious Belief*. London: Rockliff, 1956.
- Thouless, R. H. *An Introduction to the Psychology of Religion*. Cambridge, U.K.: Cambridge University Press, 1923; reprinted, 1936.

### RELEVANT THEOLOGICAL WORKS

- Raven, C. E. *Natural Religion and Christian Theology*. Vol. II. Cambridge, U.K.: Cambridge University Press, 1953.
- Schleiermacher, Friedrich. *Der christliche Glaube*. Berlin, 1821–1822. 2nd ed. translated by H. R. Mackintosh and J. S. Stewart as *Christian Faith*. Edinburgh: T. and T. Clark, 1928.
- Schleiermacher, Friedrich. *Die Religion. Reden an die Gebildeten unter ihren Verächtern*. Berlin, 1799. Translated by John Oman as *On Religion: Speeches to Its Cultured Despisers*. London: K. Paul, Trench, Trubner, 1893; reprinted, Gloucester, MA, 1958; paperback ed., New York: Harper, 1958.
- Tennant, F. R. *Philosophical Theology*. Vol. I. Cambridge, U.K.: Cambridge University Press, 1928.

### PHILOSOPHICAL STUDIES

The following works explore the relation between religious experience and religious belief or contain relevant ideas and arguments:

- Buber, Martin. *Ich und Du*. Leipzig, 1923. Translated by R. Gregor Smith as *I and Thou*. Edinburgh, 1937.
- Campbell, C. A. *On Selfhood and Godhood*. London: Allen and Unwin, 1957. On Rudolf Otto see Lecture 16.
- Hick, John, ed. *Faith and the Philosophers*. New York: St. Martin's Press, 1964.
- Holland, R. F., and J. M. Cameron. "Religious Discourse and Theological Discourse." *Australasian Journal of Philosophy* 34 (1956): 147–163, 203–207.
- Hook, Sidney, ed. *Religious Experience and Truth*. New York: New York University Press, 1961.
- Lewis, H. D. *Our Experience of God*. London: Allen and Unwin, 1959.
- Lewis, H. D., and C. H. Whiteley. "The Cognitive Factor in Religious Experience." *PAS*, Supp. 29 (1955): 59–92.
- Martin, C. B. *Religious Belief*. Ithaca, NY: Cornell University Press, 1959.
- Otto, Rudolf. *Das Heilige*. Breslau: Trewendt and Granier, 1917. Translated by John W. Harvey as *The Idea of the Holy*, 2nd ed. London, 1950.
- Smart, Ninian. *Reasons and Faiths*. London: Routledge and Paul, 1958.
- Smith, N. Kemp. "Is Divine Existence Credible?" *Proceedings of the British Academy* 17 (1931).

### OTHER RECOMMENDED TITLES

- Alston, William. *Perceiving God: The Epistemology of Religious Experience*. Ithaca, NY: Cornell University Press, 1991.
- Bagger, Matthew. *Religious Experience, Justification, and History*. Cambridge, U.K.: Cambridge University Press, 1999.
- Clouser, Roy. *Knowing with the Heart: Religious Experience and Belief in God*. Downers Grove, IL: InterVarsity Press, 1999.
- Davis, Caroline. *The Evidential Force of Religious Experience*. New York: Oxford University Press, 1989.
- Ellwood, Robert. *Mysticism and Religion*. Englewood Cliffs, NJ: Prentice Hall, 1980.
- Forsthoefel, Thomas. *Knowing beyond Knowledge: Epistemologies of Religious Experience in Classical and Modern Advaita*. Burlington: Ashgate, 2002.
- Gellman, Jerome. *Experience of God and the Rationality of Theistic Belief*. Ithaca, NY: Cornell University Press, 1997.
- Griffith-Dickson, Gwen. *Human and Divine: An Introduction to the Philosophy of Religious Experience*. London: Duckworth, 2000.
- Gutting, Gary. *Religious Belief and Religious Skepticism*. Notre Dame: University of Notre Dame Press, 1982.
- Mackie, J. L. *The Miracle of Theism*. Oxford: Clarendon Press, 1982.
- Miles, Thomas. *Religious Experience*. London: Macmillan, 1972.
- Pine, Nelson. *Mystic Union: An Essay in the Phenomenology of Mysticism*. Ithaca, NY: Cornell University Press, 1992.
- Proudfoot, Wayne. *Religious Experience*. Berkeley: University of California Press, 1985.
- Rowe, William. "Religious Experience and the Principle of Credulity." *International Journal for Philosophy of Religion* 13 (1982): 85–92.
- Smith, John. *Experience and God*. New York: Oxford University Press, 1968.
- Swinburne, Richard. *The Existence of God*. Oxford: Clarendon Press, 1979.

Wainwright, William. *Mysticism*. Brighton: Harvester Press, 1981.

Yandell, Keith. *The Epistemology of Religious Experience*. Cambridge: Cambridge University Press, 1993.

**Ronald W. Hepburn (1967)**  
*Bibliography updated by Christian B. Miller (2005)*

## RELIGIOUS INSTINCT

See *Common Consent Arguments for the Existence of God*

## RELIGIOUS LANGUAGE

Utterances made in religious contexts are of many sorts. In the performance of public and private worship men engage in acts of praise, petition, thanks, confession, and exhortation. In sacred writings we find historical records, dramatic narratives, proclamations of law, predictions, admonitions, evaluations, cosmological speculations, and theological pronouncements. In devotional literature there are rules of conduct, biographical narratives, and introspective descriptions of religious experience. Philosophical discussions of religious language have concentrated on a restricted segment of this enormous diversity, namely, theological statements, that is, assertions of the existence, nature, and doings of supernatural personal beings.

There are two reasons for this emphasis. First, the crucial problems about religious language appear in their purest form in theological statements. If we consider a petitionary prayer or a confession, what is puzzling about it is not the act of petition or confession, but the idea of addressing it to God, and God answering it. It is the concept of communication with a supernatural incorporeal person that seems unclear. And this lack of clarity is most apparent in the statement that there exists a God who communicates with men in various ways. We may say that the difficulties in understanding other forms of religious language all stem from obscurities in statements about God.

The second reason for philosophical concentration on theological statements lies in the fact that the philosophy of religion is primarily concerned with questions of justifiability, significance, and value. And it has generally been supposed that whether religion is a justifiable form of human activity largely depends on whether there are sufficient grounds for accepting the theological statements on which it is based. Christianity is a justifiable

institution if and only if we are warranted in accepting the proposition that the world is created and governed by an omnipotent, perfectly good personal deity who has revealed himself to men in the Bible. Thus the philosophy of religion is largely taken up with examining the grounds of religious statements. And it is when we do this that we become most acutely aware of the puzzling aspects of religious language. When we make a determined effort to decide whether it is true that God created the physical universe, it is difficult to avoid realizing how unclear what we are saying is, what implications it has, what it logically excludes, and what would count for or against it. Thus the philosophical investigation of religious language focuses on those indeterminacies in theological statements that hamper attempts to find rational grounds for acceptance or rejection.

### MEANING OF THEOLOGICAL PREDICATES

Most philosophers who have concerned themselves with the problem have located the difficulties of religious language in the predicates of theological statements. (What does “good” mean in “God is good”?) It may seem that we should start with the subject of the statement, with the concept of God. But there is really no alternative to starting with the predicates. For the only way to make clear what one means by “God” is to provide an identifying description, such as “the creator of the universe”; and to understand that phrase one must understand the predicate “created the universe” as applied to God. Theological predicates can be divided into negative (infinite, nontemporal, incorporeal) and positive. The positive predicates can be concerned either with attributes (good, wise, omniscient) or with actions (makes, forgives, speaks, watches over). Negative predicates present no special difficulty, but in themselves they are clearly insufficient to give any positive conception of the deity. Of the positive attributes we shall concentrate on attributions of action, partly because action terms pose more severe problems, partly because other attributes are dependent on them. (To say that God is wise is to say that he acts wisely; if we cannot understand what it is for him to perform one or another action, we cannot understand the attribution of wisdom to him.)

**DERIVATION AND APPLICATION.** When one reflects on the use of predicates in theological statements one comes to realize two fundamental facts: (1) this use is necessarily derivative from the application of the predicates to human beings and other observable entities; (2)

the theological use of predicates is markedly different from the application of predicates to human beings.

Theological predicates are derivative primarily because it is impossible to teach theological language from scratch. How would one teach a child what it means to say "God has spoken to me" without first making sure that the child knows what it is for a human being to speak to him? In order to do so one would have to have some reliable way of determining *when* God was speaking to him, so that when this happens one could say to him, "That is what it is for God to speak to you." And even if we admit that God does speak to people from time to time, there is no way for one person to tell when God is speaking to another person unless the other person tells him, which would require that the other person have already mastered the theological use of language. Hence there is no alternative to the usual procedure of teaching the theological use of terms by extension from their application to empirically observable objects.

As for the difference in the use of predicates as applied to God and to human beings, there are many ways of seeing that the terms cannot have quite the same meaning in both cases. If, as in classical Christian theology, God is conceived of as not in time, then it is clear that God's performance of actions like speaking, making, or comforting is something radically different from the temporally sequential performance of actions by human beings. St. Thomas Aquinas in his famous discussion of this problem based the distinction between the application of predicates to human beings and the application of predicates to God on the principle that God is an absolute unity and that, therefore, various attributes and activities are not distinguishable in God as they are in men. But even if we allow God to be temporal and straightforwardly multifaceted, we are left with the corporeal-incorporeal difference. If God does not have a body, it is clear that speaking, making, or comforting cannot be the same thing for God as for man.

This leaves us with a serious problem. We must show how the theological use of these terms is derived from their nontheological use. Until we do, it will be unclear just what we are saying about God in such utterances. The usual way of dealing with this problem is by cutting out the inapplicable portions of the original meaning of the terms, leaving the remainder for theology. Thus, since God is incorporeal, his speaking cannot involve producing sounds by expelling air over vocal cords. What is left is that God does something that results in the addressee having an experience of the sort he would have if some human being were speaking to him. The nature of the

"something" is deliberately left vague. Since God is a pure spirit, it will presumably be some conscious mental act; perhaps an act of will to the effect that the addressee shall have the experience of being told such-and-such. More generally, to attribute any interpersonal action to God is to attribute to him a purely mental act that has as its intended result a certain experience, like the one that would result from such an action on the part of a human being.

This account may throw some light on the content of statements about God, but religious thinkers have become increasingly dissatisfied with it. For one thing, it represents theological statements as metaphysical speculations and does little to illuminate the ways they fit into religious activity. Having postulated a pure immaterial substance performing mental acts that, miraculously, have effects in human experience, how do we go about getting into communication with this immaterial substance? Why should it be worshiped at all, and if it should, why in one way rather than another? Moreover, this line of reasoning is not helpful in our efforts to verify theological statements. It offers no hints on how we might determine whether our statements are true, or even whether there is such a being that performs the actions in question.

## VERIFIABILITY OF THEOLOGICAL STATEMENTS

Recent discussions have concentrated on the problem of verifiability. In the last few decades a great many philosophers have come to accept some form of the "verifiability theory of meaning," according to which one is making a genuine factual assertion, a real claim as to the way the world is, only if it is possible to conceive of some way in which what he is saying can be shown to be true or false by empirical observation. Applying this theory to theology, it has been argued that since an empirical test is in principle impossible to carry out for statements about a supernatural incorporeal personal deity, these statements cannot be regarded as straightforward factual assertions, but must be interpreted in some other way.

John Wisdom in his influential essay, "Gods," analogizes the function of theology to the following situation. Two people return to a long-neglected garden and find some of the old flowers still surviving among the weeds. One suggests that some gardener has been caring for the plot, and the other expresses doubt about this. On investigation, it turns out that no one in the vicinity has ever noticed anyone working on the garden. Moreover they discover that gardens left to their own devices often take



this form. But the first man does not abandon his hypothesis. Instead he expresses his belief that someone who is not discernible by the senses comes and cares for the garden, carrying out designs he and his companion do not fully grasp. At this point the first man has modified his “gardener” hypothesis to the point at which it is no longer susceptible to empirical confirmation or refutation. No matter what is or is not discovered empirically, he will continue to hold it. In this case it seems plausible to say that he is no longer expressing a belief about actual objective events. If he were, he would be able to imagine, however inadequately, some way in which the existence or nonexistence of these events would be revealed to our experience. He is, rather, expressing a “picture preference.” It is rewarding to him to think of the situation as if a gardener were coming to take care of the flowers. If beliefs about God are equally refractory to empirical test, it would seem to follow that they too must be interpreted otherwise than as straightforward matters of fact. (Wisdom, however, does not commit himself to this conclusion.)

In considering the “verificationist” challenge to theology, we must scrutinize both premises of the argument: (1) theological statements are not susceptible to empirical test; (2) if they are not empirically testable they cannot be construed as factual assertions that can be assessed as true or false.

**ARE THEY EMPIRICALLY TESTABLE?** The question of whether theological statements are subject to empirical test is quite complicated. If we rule out mystical experience as a means of observation, then it is clear that statements about God cannot be tested directly. But science is full of hypotheses about unobservable entities—electromagnetic fields, social structures, instincts—which verificationists accept as meaningful because they can be tested indirectly. That is, from these hypotheses we can draw implications that can themselves be tested by observation. The question is whether directly testable consequences can be drawn from theological statements. We can phrase this question as follows: Would we expect any possible observations to differ according to whether there is or is not a God? It would clearly be unreasonable to require of the theologian that he specify a set of observations that would conclusively prove or disprove his assertions. Few, if any, scientific hypotheses could meet that requirement. The most that could reasonably be demanded is that he specify some observable states of affairs that would count for or against his assertions.

One thing that makes this problem difficult is the fact that on this point religious belief differs at different times and places. Supernatural deities have often been thought of as dealing in a fairly predictable way with contingencies in the natural world and human society. Thus in many primitive religions it is believed that the gods will bring abundant crops or victory in battle if they are approached in certain ways through prayer and ritual. Even in as advanced a religious tradition as the Judeo-Christian, it is believed that God has certain fixed intentions that will result in prayers being answered (when made in the right spirit and under proper conditions) and will result in the final victory of the church on earth.

It would seem that such expectations provide a basis for empirical test. Insofar as they are fulfilled, the theology is confirmed; insofar as they are frustrated it is disproved. However, things are not that simple. Even in primitive communities such tests are rarely allowed to be decisive; the empirical implications are hedged around with a variety of escape clauses. If the ritual dances are held and still the crops fail, there are several alternatives to abandoning traditional beliefs about the gods. Perhaps there was an unnoticed slip somewhere in the ritual; perhaps devils were conducting counterrituals. More sophisticated explanations are employed in the more advanced religions. For example, God will answer prayers, but only when doing so would be for the true good of the supplicant.

Moreover, as science develops, religion comes to be more concerned with the personal life of the worshiper and less concerned with prediction and control of the course of events. Among religious intellectuals today such predictions as are still made are clearly not testable in practice, because of their lack of specificity (“all things will work together for the good for those who love God”), their enormous scope (“everything in the world contributes to the development of moral personality”), or their inaccessibility (“after death we shall see God face to face”). Nevertheless, it seems that within religion there are strong barriers to completely divorcing belief in God from the expectation of one event rather than another; and so long as there is some connection of belief with testable predictions, however tenuous, it would be a mistake to think of religious statements as absolutely unverifiable in principle.

**ARE THEY ASSERTIONS OF FACT?** As to whether a statement that cannot be empirically tested must not be construed as an assertion of fact, a theologian might well challenge the application of the verifiability theory to the-

ology. If God is supernatural, we should not expect his behavior to be governed by any laws or regularities we could hope to discover. But then we could never be certain that, for example, the statement that God loves his creatures would ever imply that a war should have one outcome rather than another. This would mean that, according to the verifiability theory, it would be impossible for us to make any statements, even false ones, about such a being. But a theory that would prevent us from recognizing the existence of a certain kind of entity, if it did exist, would be an unreasonable theory.

### NONASSERTIVE INTERPRETATIONS

Be that as it may, a number of philosophers have been so impressed by these difficulties over verifiability that they have tried to construe theological utterances as something other than straightforward factual assertions. Attachment to the verifiability theory is not the only motivation behind the development of such theories. There are those, like George Santayana, who, without holding that theological sentences are factually meaningless, are convinced that as factual assertions they are false, but still are unwilling to abandon traditional religious discourse. They feel that somehow it has a valuable function in human life, and in order to preserve it they are forced to reinterpret it so that the unwarranted factual claims are expunged. Still another motivation is the hope that this will contribute to the resolution of the problem mentioned earlier, that of specifying the way predicates are used when they are applied to God. As we saw, attempts to give an illuminating definition of theological predicates have not been wholly successful, and this can be taken to indicate that a different sort of approach is needed.

One such line of investigation takes sentences as its units rather than words. It focuses on the kind of linguistic act performed when theological sentences are uttered, rather than on the meaning of words in theological contexts. Instead of asking what “forgives” means when applied to God, we ask what linguistic action is performed when one uses the sentence “God forgives the sins of those who truly turn unto him.” It is this sort of question one is asking when one wonders whether theological sentences make factual assertions and, if not, what they are used to do. If we could answer this question we would have made sufficiently clear how words are being used in theological sentences without having to define special senses for constituent words.

Nonassertive interpretations can be divided into four groups. Statements about God have been interpreted as

(1) expressions of feelings of various sorts; (2) symbolic presentations of a variety of vital aspects of experience, from natural facts to moral ideals; (3) integral elements in ritualistic worship; (4) a unique kind of “mythical” or “symbolic” expression, not reducible to any other use of language.

**EXPRESSIONS OF FEELING.** Theological utterances have been interpreted as expressions of feelings that arise in connection with religious belief and activity. Thus we might think of “God made the heavens and the earth” as an expression of the sense of awe and mystery evoked by grandeur of nature; of “God has predestined every man to salvation or damnation” as an expression of a pervasive sense of helplessness; and of “God watches over the affairs of men” as an expression of a sense of peace, security, at-homeness in the world. This is “poetic” expression rather than expression by expletives. It is like expressing a sense of futility by saying “life’s a walking shadow” rather than like expressing futility by saying “Ah, me.” That is, the feeling is expressed by depicting a situation that might naturally evoke it; a sense of security, for instance, is evoked by some powerful person looking after one.

**SYMBOLIC PRESENTATIONS.** Symbolic interpretations of religious doctrines have been common for a long time. The story of Noah and the Flood has been regarded by many Christian thinkers not as an account of actual historical occurrences, but rather as a symbolic way of presenting certain religiously important points—that God will punish the wicked, but will also, under certain conditions, show mercy. Many of the traditional ways of speaking about God have to be taken as symbolic. God cannot literally be a shepherd or a rock. The shepherd functions as a symbol of providence and the rock as a symbol for God’s role as a refuge and protection in time of trouble. A symbol in this sense is some (relatively) concrete object, situation, or activity that can be taken to stand for the ultimate object of discourse through some kind of association, usually on the basis of similarity. We speak symbolically when what we literally refer to is something that functions as a symbol.

In the traditional use of symbolic interpretation it is, necessarily, only a part of theological discourse that is taken as symbolic. For if we are to hold that the symbolic utterances are symbolizing facts about God, we will have to have some way of saying what those facts are; and we cannot make that specification in symbolic terms, on pain of an infinite regress. But we are now considering views according to which all theological discourse is symbolic, which means that if we are to say what is being

symbolized it will have to be something in the natural world that can be specified in nontheological terms. The most common version of such a view is that theological utterances are symbolic presentations of moral ideals, attitudes, or values. This position has been set forth most fully and persuasively by George Santayana, and in a more up-to-date form by R. B. Braithwaite. According to Santayana every religious doctrine involves two components: a kernel of moral or valuational insight, and a poetic or pictorial rendering of it. Thus the doctrine that the physical universe is the creation of a supremely good personal deity is a pictorial rendering of the insight that everything in the world is potentially usable for the enrichment of human life. The Christian story of the incarnation, sacrificial death, and resurrection of Jesus Christ is a way of making the point that self-sacrifice for others is of supreme moral value. It is worthwhile embodying these moral insights in theological doctrine because this vivid presentation, together with the systematic cultivation of feelings and attitudes that accompanies it, provides a more effective way of getting across the insights than would a bald statement.

The way in which interpretations of the first two kinds throw light on the theological use of predicates is analogous to the way in which one explicates the use of words in poetic metaphors. If we consider the metaphor in “sleep that knits up the raveled sleeve of care,” it is clear that “knit” is not used simply to refer to a certain kind of physical operation. This utterance has quite different kinds of implications from “she knit me a sweater,” in which *knit* does have its usual sense. In the metaphoric statement, *knit* is used in its usual sense to depict a certain kind of situation that, as a whole, is presented as an analogue of the effect of sleep on care. The only way of effectively getting at the function of the word *knit* is by seeing how the whole phrase “knits up the raveled sleeve” is used to say something indirectly about sleep.

In the first two of the four kinds of nonassertive interpretation we are examining, theological statements are essentially metaphors. And if they are correctly so regarded, we get nowhere if we extract the word *made* from the sentence “God made the heavens and the earth” and try to say what it means by itself. What we have to do is take the picture presented by the whole sentence and see how it functions as a way of expressing a feeling of security, or as a way of presenting the insight that everything in the world can be used to enrich human life.

**RITUALISTIC INTERPRETATION.** The ritualistic interpretation of theological discourse can best be introduced

by citing the reply of an intellectually sophisticated high-church Anglican to a question from an agnostic friend. The question was, “How can you go to church and say all those things in the creed?” The reply: “I don’t say them; I sing them.” In the view under consideration, the corporate practice of worship is the native soil from which talk about God springs. Talk about the attributes, doings, and intentions of a supernatural personal being has meaning as a part of the practice of worship and is puzzling only when it is separated from that context. If we think of an utterance like “God made the heavens and the earth” as the expression of a belief about the way things in fact originated and then wonder whether it is true or false, we will be at a loss. To understand it we have to put it back into the setting where it (or rather a second-person correlate, such as “Thou, who hast made the heavens and the earth”) does its work. In that setting, these words are not being used to explain anything, but to do something quite different.

Unfortunately, proponents of this view have never been very clear about what this “something different” is. The clearest suggestion they give is that the talk about God serves to provide an imaginative framework for the conduct of worship. It articulates one’s sense that something important is going on, and it helps to indicate the appropriateness of one response rather than another. In speaking of the sacrament of communion as the reenactment of the self-sacrifice of an omnipotent personal deity who took on human form, and in conceiving of it as a cleansing and renewing incorporation of the substance of such a deity, one provides for the activity a pictorial framework that records and nurtures the felt solemnity of the occasion and the attitudes and aspirations kindled by the ceremony. This position presupposes, contrary to the usual view, that ritual worship has an autonomous value, apart from any theological foundation. It is generally supposed that a given ritual has a point only if certain theological doctrines are objectively true. But in the ritualistic interpretation, theological doctrines are not regarded as statements about which questions of truth or falsity are properly raised. Since these doctrines depend for their significance on the ritual, it is supposed that the ritual has some intrinsic value in forming and giving expression to valuable sentiments, feelings, and attitudes.

**MYTHS.** Ernst Cassirer has developed the notion that the basis of religious discourse lies in a unique “symbolic form” that he terms “mythical.” He maintains that it is found in purest form in the myths of primitive peoples and is based on a way of perceiving and thinking about the world that is radically different from our accustomed

mode. In the “mythical consciousness” there is no sharp distinction between the subjective and the objective. No clear line is drawn between symbol and object, between wish and fulfillment, between perception and fantasy. Again, no sharp distinction is made between the object itself and the emotional reaction it evokes; emotional response is taken to be an integral part of the environment. As a result none of our familiar standards of truth or objectivity are applicable. What is most real is what arouses the greatest intensity of emotional response and, particularly, what is felt as most sacred. (The sacred-profane distinction is the fundamental contrast.) The mythical consciousness carries its own special organizations of space and time. For example, there is no distinction made between a position and what occupies it; every spatial position is endowed with a qualitative character and exerts influence as such.

It is the view of Cassirer, and of followers such as Susanne Langer, that sophisticated theology represents an uneasy compromise between mythical and scientific modes of thought, and as such cannot be understood without seeing how it has developed from its origins. It is basically a mythical view of the world, given a “secondary elaboration” in a vain attempt to make it acceptable to the rationalistic consciousness; judged by rationalistic standards it is not only groundless, but meaningless.

*Mysticism.* Philosophers and theologians in the mystical tradition have put forward versions of this fourth kind of interpretation that do not regard theology as a manifestation of cultural lag. To the mystic the only way to communicate with God is through mystical experience, and this experience reveals God to be an ineffable unity. He can be directly intuited in mystical experience, but since there are no distinctions within the absolute unity of his being, and since any statement we can make predicates of him one thing rather than another, for example, wisdom as distinguishable from power, no statement can be true of him. The most we can do in language is to direct our hearers to the mode of experience that constitutes the sole means of access. Proponents of this view sometimes speak of theological language as “symbolic,” but this differs from our second type of theory in that here there is no way to make explicit what it is that the theological utterances symbolize, and it is therefore questionable whether we should use the term *symbol*. A symbol is always a symbol of something. In fact it is difficult to make clear just what, on this view, religious utterances are supposed to be doing. They are said to “point to,” “adumbrate,” or “indicate” the ineffable divine reality, but all too often these expressions remain uninterpreted.

In recent years two interesting attempts have been made to develop this position further. W. T. Stace, in his book *Time and Eternity* (1952), considers the chief function of religious language to be the evocation of mystical experience, or faint echoes thereof. This seems at first to be a subjectivist account, with the deity omitted, but, as Stace correctly points out, it is an axiom in the mystical tradition that no difference can be found in mystical experience between subject and object, and on these grounds Stace refuses to make the distinction. Although Stace goes along with the mystical tradition in regarding mystical experience as ineffable, he departs from this official position to the extent of giving some indications of the aspects of this experience that different theological utterances evoke. “God is truth” evokes the sense of revelatoriness, “God is infinite” the sense of all-inclusiveness, “God is love” the blissful, rapturous character of the experience, and “God is one” the absolute unity of the experience and the sense of the dissolution of all distinctions.

Paul Tillich, although not squarely in the mystical tradition, is faced with similar problems in the interpretation of religious language. He holds that theological doctrines “symbolize” an ultimate reality, “being-itself,” about which nothing can be said literally except that it is metaphysically ultimate. In attempting to clarify the function of religious language, Tillich develops the notion that it is an expression of “ultimate concern,” a complex of devotion, commitment, and orientation, focused on something nonultimate—a human being, a nation, or a supernatural deity. Religious statements, which literally refer to such relatively concrete focuses of ultimate concern, express the sense of the sacredness such objects have as “manifestations” of being-itself. But just what it is for such an object to be taken as a “manifestation” or “symbol” of being-itself, Tillich never makes clear.

The basic weakness in these mythical and mystical interpretations is the failure to present any clear hypothesis concerning the function of religious language. Even Cassirer’s ideas on “mythical thought” have never been developed to the point of clarifying what contemporary religious believers mean when they talk about God. The other positions are more intelligible, and they all base themselves on important aspects of the use of language in religion. But it seems that each, by inflating its chosen aspect to sole authority, has killed the goose that lays the golden eggs. There is no doubt that in talking about God, religious people express feelings of various sorts, present moral ideals, and articulate what is going on in ritual. But it is not at all clear that they would be using this kind of

language if they were not convinced of the truth of the statements they make. Why should I express a feeling of security by saying “God made the heavens and the earth” unless I believe, or at least have some tendency to believe, that as a matter of objective fact the physical universe owes its existence to the creative activity of a supernatural personal deity? Still more, why should I take on the complex of attitudes and activities that goes along with this assertion unless I believe it to be true?

The statement-making function is the cornerstone on which all the other functions depend. And if one is convinced that theological statements are either false or meaningless and still wants to hold to traditional religious formulations, one may *propose* a reinterpretation of theological utterances as expressions of feeling or symbolizations of natural facts. But a proposal for adopting a certain interpretation must be distinguished from a claim that the proposed interpretation correctly reflects the way doctrines are commonly understood.

It would seem that talk about God is much more complex than is recognized by any of the existing theories. The brief discussion given above of empirically testable implications illustrates this point. Theological sentences perform a great many closely interrelated linguistic functions. In saying “God, who created the world, watches over the affairs of men,” the believer is committing himself to a certain general view of the ultimate basis of the world, giving voice to certain, perhaps very indefinitely specified, expectations as to how things will ultimately turn out, expressing a basic sense of security in life, committing himself to approach God in prayer and ritual in one way rather than another. And these functions are intimately dependent on each other. What is needed is a description of the relationships among these functions, one sufficiently complex to match the complexity of the subject matter.

**See also** Braithwaite, Richard Bevan; Cassirer, Ernst; Mysticism, History of; Philosophy of Religion, Problems of; Propositions, Judgments, Sentences, and Statements; Santayana, George; Stace, Walter Terence; Subject and Predicate; Tillich, Paul; Verifiability Principle; Wisdom, (Arthur) John Terence Dibben.

### Bibliography

St. Thomas Aquinas’s historically important discussion of the “analogical” character of theological terms is found in Question XIII of Part I of the *Summa Theologiae*. For further discussions in the Thomist tradition see Cajetan, *The Analogy of Names*, translated by E. A. Bushinski and H. J. Koren (Pittsburgh: Duquesne University, 1953), and E. L.

Mascall, *Existence and Analogy* (London, 1949). Recent discussions by practitioners of analytical philosophy are to be found in Antony Flew and Alasdair MacIntyre, eds., *New Essays in Philosophical Theology* (New York: Macmillan, 1964), and in Basil Mitchell, ed., *Faith and Logic* (London: Allen and Unwin, 1957). See also Alasdair MacIntyre, “The Logical Status of Religious Belief,” in *Metaphysical Beliefs* (London: SCM Press, 1957); John Wisdom, “Gods,” in *Essays in Logic and Language*, edited by Antony Flew, first series (Oxford, 1951); C. B. Martin, *Religious Belief* (Ithaca, NY: Cornell University Press, 1959); and I. T. Ramsey, *Religious Language* (London: SCM Press, 1957). Various kinds of symbolic interpretations of religious statements are presented in George Santayana, *Reason in Religion* (New York: Scribners, 1905); R. B. Braithwaite, *An Empiricist’s View of the Nature of Religious Belief* (Cambridge, U.K., 1955); W. M. Urban, *Language and Reality* (New York, 1939); Edwin Bevan, *Symbolism and Belief* (Boston: Beacon Press, 1957); and Philip Wheelwright, *The Burning Fountain* (Bloomington: Indiana University Press, 1954). Ernst Cassirer’s views on mythical language are set forth most completely in Vol. II of *Philosophie der symbolischen Formen*, 3 vols. (Berlin: Cassirer, 1923, 1925, 1929), translated as *The Philosophy of Symbolic Forms* by Ralph Manheim (New Haven, CT: Yale University Press, 1952, 1955, 1957); and in more concentrated form in *Sprache und Mythos* (Leipzig: Teubner, 1925), translated as *Language and Myth* by S. K. Langer (New York: Harper, 1946). For other versions of the view that religious language constitutes an autonomous mode of discourse, see W. T. Stace’s *Time and Eternity* (Princeton, NJ: Princeton University Press, 1952) and Paul Tillich’s *Systematic Theology* (Chicago: University of Chicago Press, 1951–1963) and *Dynamics of Faith* (New York: Harper, 1957).

William P. Alston (1967)

## RELIGIOUS LANGUAGE [ADDENDUM]

Two significant contributions to recent discussions of religious language are offered by Janet Soskice and William P. Alston. In *Metaphor and Religious Language* (1985) Soskice offers as a working definition “metaphor is that figure of speech whereby we speak about one thing in terms which are seen to be suggestive of another” (p.15). The minimum unit in which a metaphor is established is semantic. A satisfactory theory of metaphors “should regard metaphors neither as a simple substitution for literal speech nor as strictly emotive. Metaphors should be treated as fully cognitive and capable of saying that which may be said in no other way. It should explain how metaphor gives us “two ideas for one,” yet do so without lapsing into a comparison theory” (p.44). Noncognitive accounts of metaphor are rejected because “we cannot conceive of emotive ‘import’ apart from a

cognitive content which elicits it” (p. 27). The “two ideas for one” feature involves a metaphor having a “unity of subject matter” that “draws upon two (or more) sets of associations, ... characteristically, by involving the consideration of a model or models” (p. 49). A model is “an object or a state of affairs ... viewed in terms of its resemblance, real or hypothetical, to some other object or state of affairs” (p. 100).

Models come in two types: paramorphic (the source and subject differ, as in billiard ball movement serving as a model for the properties of gases); or homeomorphic, where the subject is the source (e.g., a dummy used to teach lifesaving skills). Models, in both theology and science, are essential to theories because they carry their explanatory force. “The fertility of a theory lies in its ability to suggest possibilities of explanation which, while not inconsistent with, are more than simply the logical extensions of mathematical formulas ... this suggestive capacity ... constitutes the fruitfulness of a theory, and gives the theory the predictive nature which is its *raison d’être*” (p. 114). We do not describe God but point to God through effects, and beyond them to him. We refer without defining. “This is the fine edge at which negative theology and positive theology meet, for the apophatic insight that we say nothing of God, but only point to Him ... this separation of referring and defining is at the very heart of metaphorical speaking...” (p. 140).

Nothing in Soskice’s account of metaphor entails that language about God must be nonliteral. The claim *All language about God is metaphorical* is not metaphorical. The idea that no metaphor can be translated into or replaced by literal terms is false. Consider Soskice’s example of an expression of hope that a soldier will be pardoned eliciting, “That’s blowing on cold coals.” “There’s no chance of that” is a literal translation. “God is a rock” seems replaceable by “God is utterly reliable.” If it is not, this is a matter of the associations of “rock” in biblical and theistic literature being multiple. It does not follow that any of the things that “rock” suggests are nonliteral. It just suggests that there are a variety of possibilities, more perhaps than we can list, each of which may be perfectly expressible without remainder in literal fashion.

A basic assumption is that no literal description can be true of God. As is typical, we are referred to certain ideas: We cannot comprehend (know all there is to know) about God; descriptions of God based on religious experience are defeasible; certainty about claims concerning God is unattainable; and it is always possible that we will have to modify our concept of God. But there are an infinite number of truths concerning a golden retriever, see-

ing the golden is defeasible, certainty about it is unavailable, and we may have to revise our concept thereof. But it is not beyond literal description. Further, God can be misdescribed (e.g., “God is a cantaloupe”), which even the most deluded of empiricist positivists presumably will recognize as false. But then what, in principle, precludes God from being correctly described?

William P. Alston’s major essays concerning religious language are collected in *Divine Nature and Human Language* (1989). In “Irreducible Metaphors in Theology” he says that “in the typical metaphorical statement the speaker is ‘building on’ the relevant meaning of the predicate term in two ways ... he is presenting the thing to which the term literally applies as a model of the subject [and] ... he has in mind one or more resemblances between model and subject and he abstracts from these resemblances what he means to be attributing to the subject” (1989, p. 23). The resemblance may be either general or specific. Everything resembles everything else in some way. Any metaphor based on this fact corresponds to a literal way of expressing the similarity. Regarding metaphors intended to express truths, he writes: “Though irreducible metaphors seem to promise a way of combining the denial of predication in theology with the preservation of significant theological truth claims, this fair promise dissipates on scrutiny like mist before the morning sun. Either the panmetaphoricist abandons the aspiration to significant truth claims or he revokes the ban on literal predication” (p. 37).

“Can We Speak Literally of God?” considers predicates that apply to personal agents (“P-predicates”) in their application to God. These include mental and action predicates. These have been understood on a private paradigm model (one knows what “depression” means by being depressed) and functionally (“being depressed” refers to a state that functions efficaciously in a causal system to yield a distinctive range of behavior). The idea of basic actions that involve no bodily movements, and of nonbasic actions that involve only mental actions that bring about effects, are both intelligible and applicable to incorporeal beings. “Literal” does not mean “empirical.”

“Functionalism and Theological Language” and “Divine and Human Action” consider functional accounts of mental concepts to argue that these concepts can apply to God. We can “form the conception of a being (a ‘system’) in which some factors depend on their relations to others for being what they are, even though there are no temporally successive processes for formation of any subjection to laws. More specifically, we are to think

of God as realizing a complex structure of attitudes, knowledge, tendencies, executive intentions, and volitions in the ‘eternal now’ ...” (p. 99). The stability of this system is to be understood, not by way of there being laws that hold regarding it, but by way of essential properties of the system. But this gives us only a description of God as a system of items that bear various dependence or causal relationships, not of a personal agent. Insofar as the relevant concepts are strictly functionalist, they do not entail even consciousness. When we turn to religious discourse about God, the functionalist account is not nearly enough: “For the religious life, we need to go beyond that in ways that launch us into the still not sufficiently charted seas of the figurative and the symbolic” (p. 103).

“Referring to God” distinguishes between direct reference and reference by description. Reference by description offers a description that is true only of the referent; direct reference names an object of one’s experience. Direct reference to God can occur only if someone experiences God (Alston takes it that some people do). Others who do not themselves experience God can then refer to the being that others have referred to; reference thus spreads throughout a religious community. Direct reference is more basic than descriptive, because if one refers to a being both descriptively and directly, and one learns that the description is false of the being directly referred to, it is the latter that determines what was the actual object of reference. Nonetheless, Alston admits that “reference could always take place via a description” (p. 107). A consequence is that it is possible that someone who thinks of God as an omnipotent, omniscient spirit, and one who thinks of God as an impersonal force, may refer to the same being. Alston says that it may be that both are “worshiping the one true God” (p. 116). If so, worship does not require much by way of actual knowledge of God.

There was never any reason to think that a causal theory of reference wedded to a functionalist account of P-predicates would yield significantly positive results regarding description of God. It seems fair to say that in spite of the sophisticated and helpful discussions provided by Soskice and Alston, accounts of religious language that are philosophically articulate and allow for seriously realistic accounts in theology remain more matters on the agenda than they are accomplishments of current work in the field.

**See also** Alston, William P.; Metaphor; Subject and Predicate.

### **Bibliography**

- Alston, William P. *Divine Nature and Human Language: Essays in Philosophical Theology*. Ithaca, NY: Cornell University Press, 1989.
- Alston, William P. *Illocutionary Acts and Sentence Meaning*. Ithaca, NY: Cornell University Press, 2000.
- Soskice, Janet M. *Metaphor and Religious Language*. Oxford: Oxford University Press, 1985.

**Keith E. Yandell (2005)**

## RELIGIOUS PLURALISM

The fact that there is a plurality of religions is significant in different ways from different points of view. From a skeptical point of view their different and often incompatible beliefs confirm the understanding of religion as delusion. Thus, Bertrand Russell wrote that “It is evident as a matter of logic that, since [the great religions of the world] disagree, not more than one of them can be true” (1957, xi). From the point of view of an exclusive and unqualified commitment to any one religion the fact of religious plurality is readily coped with by holding that all religions other than one’s own are false, or false insofar as their belief systems differ from one’s own. But from a point of view that sees religion as a worldwide phenomenon that is not to be dismissed in toto as delusion but as the human response to a divine/transcendent/ultimate reality, the fact of plurality poses a major philosophical problem. On the one hand, the “great world religions” seem—to many impartial observers, at any rate—to affect human life for both good and ill to more or less the same extent. But on the other hand their respective belief systems, although having important similarities, also include starkly incompatible elements. According to some the Real (a term at home in the Judeo-Christian tradition and corresponding to the Sanskrit *sat* and the Arabic *al-Haqq*) is personal but according to others not personal. And within each group of religions there are wide differences. Is the ultimate Person the Christian Trinity or the Qur’anic Allah, or the Adonai of Judaism, or Vishnu, or Shiva? Is the nonpersonal Ultimate the Brahman of advaitic Hinduism, or the Dao, or the Dharmakaya or Void or Nirvāṇa of the Buddhist traditions? And how could the Real be all of these at once? The logic of religious difference here is in fact very complex, as is shown by William Christian’s analysis (1987).

The problem is particularly acute for a major form of religious apologetic that became prominent in the 1980s and 1990s. This holds that the basic empiricist principle that it is rational, in the absence of specific overriding

considerations, to base beliefs on experience should be applied impartially to all forms of putatively cognitive experience, including religious experience—unless, again, there are specific overriding considerations to the contrary. This has been argued directly by William Alston (1991) and others and indirectly by Alvin Plantinga (in Plantinga and Wolterstorff 1983), whose defense of the rationality of holding “properly basic” religious beliefs presupposes religious experience as their ground.

Most of the philosophers who employ this kind of apologetic have applied it only to specifically Christian beliefs. But it is evident that precisely the same argument is available for the belief systems of other religions. If Christian religious experience renders it epistemically justifiable (subject to the possibility of specific reasons to the contrary) to hold Christian beliefs, then Buddhist religious experience renders it epistemically justifiable, with the same qualification, to hold Buddhist beliefs, Muslim religious experience to hold Muslim beliefs, and so on. Thus, anyone who maintains that the Christian belief system is true, but that the belief systems of Buddhists, Muslims, and so on are false insofar as they differ from it, has implicitly reversed the original apologetic and is presenting Christian religious experience as the sole exception to the general rule that religious experience gives rise to false beliefs!

Alston, recognizing the challenge posed by the fact of religious diversity to the experiential apologetic, has responded by saying that in this situation it is proper for the Christian to continue within her own belief system, despite the existence of other equally well-justified alternatives, while, however, she seeks “a way to show in a non-circular way which of the contenders is correct” (1991, p. 278).

An alternative use of the experiential apologetic rejects the assumption that only one of the different religious belief systems can be true. This approach (Hick 1989) distinguishes between, on the one hand, the ultimate religious reality, the Real, beyond the scope of our (other than purely formal) human conceptualities, and, on the other hand, the range of ways in which that reality is humanly conceived, and therefore humanly experienced, and therefore humanly responded to within the different religiocultural ways of being human. The epistemology operating here is one that, in the Kantian tradition, recognizes an important contribution by the perceiver to the form a reality is perceived to have. As Thomas Aquinas wrote, “Things known are in the knower according to the mode of the knower” (*Summa Theologiae*, II/II, 1, 2). And in religious knowing the mode of the

knower differs from religion to religion. From this point of view the fact of religious diversity does not constitute a challenge to the experiential apologetic but rather a series of examples of its valid application.

Other philosophical responses to the fact of religious plurality, not specifically related to the experiential apologetic, include the “perennial philosophy” (e.g. Schuon 1975, Smith 1976), which distinguishes between the essence (or esoteric core) of religion and its accidental (or exoteric) historical forms. In their esoteric essence all the great traditions converge in a transcendental unity, the Absolute Unity that is called God. Experientially, this sees the mystics of the different religions as participating in an identical experience, although they articulate it in the different ways provided by their traditions. This view is opposed by those (e.g., Katz 1978) who hold that all experience is concept laden and that mystical experience accordingly takes different forms within the different traditions.

There is also the view of John Cobb (in Kellenberger 1993) that the religions are directed toward different ultimates, particularly the personal reality worshiped in the theistic religions and the nonpersonal process of the universe experienced in Buddhism. Yet other constructive suggestions include those of Joseph Runzo (1986), James Kellenberger (1989), and the authors included in the symposium *Inter-Religious Models and Criteria* (Kellenberger 1993).

**See also** Philosophy of Religion; Philosophy of Religion, History of; Philosophy of Religion, Problems of; Religious Experience; Russell, Bertrand Arthur William; Thomas Aquinas, St.

### **Bibliography**

- Alston, W. P. *Perceiving God*. Ithaca, NY: Cornell University Press, 1991.
- Basinger, David. *Religious Diversity: A Philosophical Assessment*. Burlington: Ashgate, 2002.
- Christian, W. A. *Doctrines of Religious Communities*. New Haven, CT: Yale University Press, 1987.
- Godlove, T. F. *Religion, Interpretation, and Diversity of Belief*. Cambridge: Cambridge University Press, 1989.
- Griffiths, Paul. *Problems of Religious Diversity*. Oxford: Blackwell, 2001.
- Hick, J. *An Interpretation of Religion*. New Haven, CT: Yale University Press, 1989.
- Katz, S. T., ed. *Mysticism and Philosophical Analysis*. London: Sheldon Press, 1978.
- Kellenberger, J. *God: Relationships with and without God*. New York: St. Martin's Press, 1989.
- Kellenberger, J., ed. *Inter-Religious Models and Criteria*. New York: St. Martin's Press, 1993.



- Krieger, D. J. *The New Universalism*. Maryknoll, NY: Orbis, 1991.
- McKim, Robert. *Religious Ambiguity and Religious Diversity*. Oxford: Oxford University Press, 2001.
- Plantinga, A., and N. Wolterstorff, eds. *Faith and Rationality*. Notre Dame, IN: University of Notre Dame Press, 1983.
- Quinn, Philip, and Kevin Meeker, eds. *The Philosophical Challenge of Religious Diversity*. New York: Oxford University Press, 2000.
- Runzo, J. *Reason, Relativism, and God*. New York: St. Martin's Press, 1986.
- Russell, B. *Why I Am Not a Christian*. London: Allen and Unwin, 1957.
- Schuon, F. *The Transcendent Unity of Religions*, rev. ed. New York: Harper and Row, 1975.
- Smith, H. *Forgotten Truth*. New York: Harper and Row, 1976.
- Vroom, H. M. *Religions and the Truth*. Amsterdam: Rodopi, 1988.

**John Hick (1996)**

*Bibliography updated by Christian B. Miller (2005)*

## RENAISSANCE

“Renaissance” is the term customarily employed to designate a cultural movement that began in Italy in the middle of the fourteenth century and spread throughout the rest of Europe. Although the term is well established in the writings of historians, its usefulness has been challenged. Indeed, there has grown up around the concept of the Renaissance an extensive controversy that sometimes threatens completely to divert the attention of scholars from the historical facts. In part, this controversy is simply an acute form of the general problem of periodization in history. The concept of the Renaissance, however, arouses particularly strong opposition because it involves a disparagement of the preceding period, the Middle Ages (*medium aevum*), from which culture presumably had to be awakened.

The idea of a rebirth of literature or of the arts originated in the period itself. Petrarch in the fourteenth century hoped to see an awakening of culture, and many later writers expressed their conviction that they were actually witnessing such an awakening in their own time. Latin was generally the language used by cultivated men to discuss such matters, but no single Latin term or phrase became the standard name for the whole cultural epoch. One of the earliest historians of philosophy in the modern sense, Johann Jakob Brucker, in 1743 referred to the Renaissance only as the “restoration of letters” (*restauratio literarum*), and wrote of the “recovery of philosophy” (*restitutio philosophiae*): Even in an earlier German work he used such Latin phrases. Scholars who wrote in Latin

never used *rinascencia* as the name for the cultural epoch as a whole. It was the French word *renaissance* that finally acquired this status and was then adopted or adapted into other languages. During the seventeenth century, and fitfully before, French scholars used the phrase *renaissance des lettres* for the humanists’ *restitutio bonarum literarum*, taking over in the process the humanist periodization of history. Other writers translated the Latin phrase or phrases into their own vernacular: Edward Gibbon (1787) spoke of the “restoration of the Greek letters in Italy,” while Heinrich Ritter, in his history of philosophy (1850), remarked that the *Wiederherstellung der Wissenschaften* derived its name from philology.

Various French authors used the term *renaissance* in titles of their works before Jules Michelet devoted one of his volumes on sixteenth-century France to *la Renaissance* (1855). However, Michelet gave only the sketchiest characterization of the period, and hardly deserves to be credited (if indeed any one person can be) with having “invented” the concept of the Renaissance. Michelet did coin one memorable phrase: He remarked that two things especially distinguished the Renaissance from previous periods—“the discovery of the world, the discovery of man.” This phrase was also used by the Swiss cultural historian Jakob Burckhardt for the title of a chapter in his famous work, *The Culture of the Renaissance in Italy* (1860). At his hands, the concept of the Renaissance received what was to become its classic formulation; all subsequent discussion of the concept invariably focuses upon Burckhardt’s description of the essential features of life during the Renaissance. Burckhardt, taking the term in its narrow sense of a literary revival of antiquity, conceded that there had been earlier “renaissances” in Europe; but he insisted that a renaissance in this sense would never have conquered the Western world had it not been united with the “already-existing spirit of the Italian people” (*italienischen Volksgeist*). Not until the time of Petrarch, so Burckhardt held, did the European spirit awake from the slumber of the Middle Ages, when the world and man lay “undiscovered.”

The relation of the Renaissance to the era that preceded it has been much studied because defenders of medieval culture quickly came to the rescue of their period, stressing its continuity with, or even its superiority to, the Renaissance. However, little has been done to clarify the relation of the Renaissance to the Enlightenment. This is rather surprising, for there was an issue that ran straight through the thought of both these eras: “Can we modern men hope to equal or even excel the achievements of antiquity?” This issue is known to literary histo-

rians as the “quarrel of the ancients and moderns.” We think of Bernard Le Bovier de Fontenelle in the seventeenth century as the main champion of the moderns, who had science and truth on their side, as against those writers, with their inflexible rules, who favored the ancients. However, much the same attitude as Fontenelle’s is found in the *De Disciplinis* of the Renaissance humanist Juan Luis Vives, who wrote in the early sixteenth century. The Renaissance itself had championed the moderns even before modern science had arisen to prove their case. Renaissance confidence in men’s powers was based on art and literature rather than on science, but it was strong nevertheless. Men could respect classical excellence and yet strive to outdo the ancients in every field, including vernacular literature.

### CHRONOLOGICAL LIMITS

Various events have been taken as marking the beginning of the Renaissance: the crowning of Petrarch as poet laureate of Rome in 1341; the short-lived triumph of Cola di Rienzi in setting up a republican Rome in 1347, an attempt to revive Rome’s former greatness; the arrival in Italy of Greek *émigrés* (which actually antedated by a few years the much publicized fall of Constantinople in 1453); the opening up of new trade routes to the East. Each choice represents the selection of a particular field as central in the history of the period: art, architecture, religion, politics, economics, trade, or learning. In certain fields it is hard to maintain any sharp break between conditions in, let us say, 1300 and those in 1350. However, few students of the history of art or of literature are prepared to deny completely the start of new trends in the fourteenth century (at least in Italy). In literature, Petrarch’s enthusiasm for Greek antiquity must surely be accepted as inaugurating, in the eyes of men in the fourteenth century, a fresh start. In painting, there is little hesitation about ascribing a similar place to Petrarch’s contemporary, Giotto; this ascription dates from the earliest attempt at a history of art, that of Giorgio Vasari (1550). No such figures can plausibly be singled out to mark new beginnings in economic or political history.

Difficulties also surround the choice of an event to mark the end of the Renaissance: the sacking of Rome in 1527, the hardening of the Counter-Reformation via the Council of Trent in 1545, the burning of Giordano Bruno in 1600, or Galileo Galilei’s setting of experimental physics on its true path around 1600—any of these might be selected. Once again, however, a periodization that is useful in one field may prove useless in another field. Generally speaking, the era from 1350 to 1600 will

include most of the developments commonly dealt with under the heading “Renaissance.”

### GEOGRAPHICAL LIMITS

The shifting locale of the Renaissance presents problems similar to those of its chronological limits. Burckhardt’s description focused exclusively on Italy; he implied that the Renaissance, after it had been taken over by the Italian *Volksgeist*, moved on to the rest of Europe. The movement to France is usually said to have resulted from the French invasion of Italy in 1515, which gave the French nobility their first glimpse of the glories of the Italian Renaissance. No comparable event can be singled out for the bringing of the Italian Renaissance to England, unless it be the return from Italy to their native land of the classical scholars William Grocyn, Thomas Linacre, and John Colet in the last decade of the fifteenth century, or perhaps Desiderius Erasmus’s arrival there about the same time. Clearly England did enjoy a renaissance, but it is not easy to fix its dates: English literary historians prefer to discuss the Elizabethan age or the age of the Tudors, thus sidestepping the question of the relation of the English Renaissance to that of the Continent. Still less clear is the coming of the Renaissance to the German lands: German historians treat the sixteenth century as the “time of the Reformation,” and tend to discuss the Renaissance chiefly in terms of its impact upon individual reformers.

The Renaissance is sometimes called the “age of adventure.” It is not at all clear, however, that the spirit behind men’s daring and adventurous actions was entirely new: The two chief incentives toward voyages of discovery, for instance, were commercial acquisitiveness and religious zeal—attitudes by no means foreign to medieval men. It was the shutting off of Venetian trade routes through the Mediterranean by the Turks that forced Europeans to search for new routes to the East, not a new desire for scientific knowledge of geography. The Spanish *conquistadores* may have thirsted for glory, but such a thirst was characteristic of medieval knights as well as of Renaissance humanists. The motives of the Franciscan missionaries were clearly religious and medieval in spirit. Moreover, in the field of domestic trade, the resurgence of economic activity in the fifteenth century that formed the basis for the cultural developments of the Renaissance was less a matter of suddenly effective acquisitiveness than of normal recovery from the slump brought about by the Black Death in 1348.

## THE NEW LEARNING

Historians may without hesitation ascribe a rebirth of classical knowledge to the Renaissance period. The discovery of old manuscripts and the invention of printing combined to make the heritage of ancient Greece and Rome available to a far wider audience. The humanists of the fourteenth and fifteenth centuries discovered and preserved many ancient texts that had been neglected for centuries. Of these perhaps the most significant from a philosophical point of view was Lucretius's *De Rerum Natura*, but many other newly discovered texts helped to enrich men's general familiarity with antiquity and to present in full view the setting in which Greek and Roman philosophy originated.

The collecting of manuscripts could be indulged in only by noblemen or well-to-do scholars, but the invention of printing made possible a broader social base for intellectual interests. With the production of vast numbers of newly discovered texts, self-education became a real possibility, as did institutional education on a broad scale. Peter Ramus in France and Philipp Melanchthon in Germany urged the educating of the people, chiefly with the idea of promoting intelligent Christian piety.

## SCIENCE

Developments in technology and science indirectly provided material for philosophical reflection. The increased use of firearms and cannon in war, for example, made necessary the mathematical study of ballistics; and the scientific work of Benedetti and Galileo drew upon the practical experience of foundries and arsenals. However, Renaissance philosophy of science still took its cue largely from Aristotle: Francis Bacon, dissatisfied with Aristotelian logic and methodology of science, found a replacement not in the actual practices of mechanics and craftsmen but in the rhetorical method derived from Aristotle and applied to the questioning of Nature.

The most spectacular and far-reaching scientific development during the Renaissance was the heliocentric theory advanced by Nicolas Copernicus, who found hints about Pythagorean cosmology in ancient works. The Copernican theory was surely the most significant revolution ever to take place in science. Far less conspicuous, but still important, were the developments in pure and applied mathematics. Modern notation (such as the use of the "equals" sign) began to be adopted, bringing with it the possibility of greater attention to logical form.

## SOCIAL VALUES

There have been many attempts, beginning with Michelet and Burckhardt, to capture *the* mind or spirit of Renaissance man. All such attempts seem doomed to failure, for they are bound to oversimplify complex social facts. We may, however, single out four sets of social ideals that were characteristic of various groups during the Renaissance.

The ideals of the feudal nobility, medieval in origin, persisted through the Renaissance among the ruling class, although they underwent considerable refinement. The rude military virtues of camp and field gave way to the graces of the court, which were set forth most admirably in Baldassare Castiglione's book *The Courtier* (1528), one of the most influential treatises on manners ever written. In Castiglione's ideal courtier we may recognize the ancestor of our "gentleman." Works of this sort are presumably also the source of the "universal man," a concept closely associated in modern minds with the Renaissance. In the heroic life idealized by the feudal tradition, love of glory and concern for one's reputation were strong social motives. The humanists' thirst for glory, which Burckhardt emphasized, merely continues this concern but applies it to the achievements of a nonwarrior class, the "knights of the pen." The urban middle class chose, as usual, to emulate the style of life of their superiors: the modern gospel of work as a *raison d'être*, shaping the whole of life, hardly existed during the Renaissance. Few social theorists extolled the virtues of commercial activity until Martin Luther stressed the sanctity of all callings, provided they benefited one's fellow men.

Religion provided the second set of ideals, which centered upon moral salvation and involved a willingness to relinquish the world and all its goods. This mood, exacerbated in some individuals by the terror of imminent death or of eternal damnation, continued unabated throughout the Renaissance; and the entire Reformation movement has been called the "last great wave of medieval mysticism." Although such a religious concern is usually associated by modern secular critics with contempt for this world and with pessimism, it is equally compatible with a cheerful resignation in the face of unavoidable misfortunes and gratefulness for such morally harmless pleasures as life affords. A genuine tension often resulted from the opposing pulls of these religious values and of secular attitudes and this-worldliness: Aristotelian philosophers as well as humanists felt this tension during the Renaissance.

A third set of ideals, that of the ancient sage (Platonic or Stoic), was consciously adopted by Renaissance

humanists as an adjunct to Christian exhortation, for many of them felt that Christians could learn much from pagan expounders of virtue. Rarely, if ever, did a humanist attempt to replace the Christian ideal altogether: Burckhardt undoubtedly overstressed the “paganism” of the humanists.

Finally, there was the ideal of a return to nature, a flight from the complexities of sophisticated urban life to pastoral pleasures. This theme has ancient antecedents in the poetry of Theocritus and Vergil, but it emerges into new prominence with Petrarch, who also stressed the benefits of solitude. Passive delight in the beauties of nature can hardly ever be totally lacking in human beings, of course, but during the Renaissance we find an interest in such activities as gardening, the collecting of strange plants and animals, and strolling through woods and fields. Petrarch’s famous excursion to the summit of Mont Ventoux turned into an occasion for Christian self-reproach, to be sure, but his letters also abound in references to his gardening and to lone promenades in the countryside near Vaucluse.

## HUMANISM

A major role in the culture of the Renaissance was played by the humanists. All sorts of people call themselves “humanists” today, but in the early days of the Renaissance the name had a clear occupational meaning. During the fourteenth century, the traditional subjects of grammar, rhetoric, and poetry had begun to be called, after a phrase of Cicero, the *studio humanitatis*. The term *umanista* was coined (on the analogy of *artista*, also a product of university slang) to designate a teacher of these subjects in Italian universities. Such studies were by no means new in the fourteenth century; in fact, the humanists were the heirs of a less ambitious but old and respectable medieval profession, that of the *dictator* or teacher of the art of letter-writing (*ars dictaminis*). The Renaissance teachers of “humanities” placed a greater emphasis on ancient models than had the *dictatores*, but their teaching had much the same; objective. Their students often became official letter-writers or speech-makers for popes and princes. Coluccio Salutati and Leonardo Bruni, two of the most influential humanists of the fifteenth century, were chancellors of Florence. The study of Greek philosophy owes much to these two men.

Renaissance humanists did not propound a distinct philosophy but took over from Cicero and Aulus Gellius the ancient ideal of a civilized and urbane way of life that could be formed through acquaintance with Greek literature. With such a program in mind, the humanists began

to concern themselves with moral and political philosophy, and this brought them into conflict with the philosophers who taught ethics or politics in the universities. The humanists regarded the Aristotelian Schoolmen as derelict in the performance of their duties, since their teaching (so the humanists claimed) made no differences in the lives of students. The scholastic teachers, in return, regarded the humanists as dilettantes and upstarts, meddling in subjects beyond their depth. The feud of humanists with philosophers began with Petrarch’s invective against the secular Aristotelians, the so-called Averroists of his day, and continued through the seventeenth century.

We still tend to see Renaissance Aristotelianism (and medieval Scholasticism as well) through the eyes of these Renaissance humanists. Their bias has crept into most histories of philosophy, largely because the first writers of histories of philosophy shared some of the humanist attitudes. One such early historian was Brucker, whose *Critical History of Philosophy* (1742–1744) has already been mentioned. Brucker presented the Renaissance as a time when human thought emerged slowly into the light (a standard metaphor) from the tiresome labyrinths of medieval Scholasticism. He divided his treatment into various sections, dealing with schools of Greek philosophy that were “restored” during the Renaissance. In spite of his scorn for “more recent Aristotelian-scholastic philosophers,” Brucker had great respect for the philosophers who followed the “genuine philosophy of Aristotle”: Pietro Pomponazzi, Simon Porta, Jacopo Zabarella, and others. Few modern historians of philosophy pay much attention to these writers. They do, however, characteristically devote lengthy sections to Paracelsus, Jakob Boehme, Robert Fludd, and other “theosophers.” According to Brucker, these theosophers “condemn all use of reason in understanding the nature of things,” and hence do not belong to the history of philosophy; he includes them only because they have commented incidentally on philosophical matters.

Whatever his own philosophical competence may have been, Brucker had one clear advantage over most later historians: He had actually read the Renaissance writers he discussed. Much of Renaissance philosophy still awaits reevaluation based upon such actual reading of texts.

The general framework of Brucker’s treatment of Renaissance philosophy remains a useful way of dealing with most of the thought of the period. The various sects of Greek philosophy were indeed “reborn” during the Renaissance; few of them escaped some sort of revival.

There was even what might be called a genuine rebirth of Aristotle, if we mean by this what Brucker probably meant: an Aristotelianism based directly upon the Greek texts rather than upon Latin or Arabic commentators.

**ARISTOTELIANISM.** It cannot be too strongly emphasized that the main stream of philosophical inquiry during the Renaissance continued to be Aristotelian. The terms employed in philosophical discussion, the problems posed, and the characteristic solutions remain, in basic outline, Aristotelian. Almost all Renaissance philosophers show the influence of their Aristotelian school training, even when they are trying most strenuously to break the shackles of that tradition. The technical terms of philosophy (such as *propositio*, *entitas*, *realis*, *materia*, *forma*, *essentia* and many others) originated or became naturalized in the Aristotelian school-tradition, and persisted even in the writings of the most daring innovators, such as Bruno. The Aristotelian tradition, for reasons already in part suggested, remains the least known and most maligned of all Renaissance schools. Elements of the critical spirit of later medieval philosophy (Scotist and Ockhamist) formed part of the school philosophy of the Spaniard Francisco Suárez and of the Scotsman John Major.

**PLATONISM.** Platonism took on new life during the Renaissance, after having been known for centuries chiefly through Aristotle's attacks on it. There was more acquaintance with Plato during the medieval period than is generally recognized, but it is still true that Marsilio Ficino's translations into Latin (first published in 1484) gave the main impetus to the spread of Plato's doctrines. Later editions of Plato often contained Ficino's translations of Proclus and Porphyry, together with his own commentaries, which were strongly colored by his Neoplatonism. Hence, the Platonism that emerged during the Renaissance cannot be distinguished easily from Neoplatonism, for it tends to be otherworldly and religious in tone. The cultural influence of Florentine Platonism emanated from the famous academy founded by Ficino in direct imitation of Plato's school. The society that grouped itself around Ficino aimed at moral improvement and resembled in character certain lay religious societies common in Italy at that time. The whole movement of natural religion was set in motion by Florentine Platonism, as was the renewed study of Pauline theology by such men as John Colet.

Florentine Platonism is well known, by name at least, to most students of the Renaissance. Much less well known is a tradition of reconciling Plato with Aristotle,

which also found expression during the period. Byzantine scholars had brought with them to Italy an old battle over the superiority of Plato or Aristotle. During the late Renaissance this battle resolved itself into a truce, with many books written to show that Plato and Aristotle agreed on fundamentals and differed only on words or nonessentials.

**STOICISM.** Only a few late Renaissance thinkers, such as Justus Lipsius and Guillaume du Vair, committed themselves explicitly to Stoicism, but the influence of Stoic philosophy may be seen at work directly and indirectly (largely via Cicero, Seneca, and the Greek commentators on Aristotle) even during the early Renaissance. Pomponazzi's rigorous moral doctrine, for example, is strongly tinged with Stoic attitudes.

**EPICUREANISM.** Rejected with horror by medieval thinkers, who saw him through the eyes of the Church Fathers, Epicurus began to be more sympathetically known as a result of humanist activity in the fifteenth century. Previous to this time, anyone who believed that the soul perished with the body was called an Epicurean, whether he held to any other Epicurean tenet or not. Now it was no longer possible to apply this label so casually. Lucretius's great poem won immediate favor because of its sturdy poetic qualities, but, until Pierre Gassendi in the seventeenth century, no one adopted the system of Epicurus in its entirety. Nevertheless, Epicurean influence prior to Gassendi's time did foster a climate less hostile to the concepts of pleasure and utility.

**SKEPTICISM.** The direct influence of philosophical skepticism in a technical sense began with the first publication of Sextus Empiricus in 1562, from which time skepticism exercised an important influence upon European thought and literature. The religious factionalism or warfare of the sixteenth century had brought about a widespread distrust of dogmatism and fanaticism on the part of such sophisticated minds as Erasmus and Michel Eyquem de Montaigne, whose writings may have contributed to the growth of that spirit of toleration usually associated with the Enlightenment.

**THE OCCULT TRADITION.** The Renaissance was immensely receptive (perhaps more so than the Middle Ages) to occult and secret lore of all kinds, especially if it claimed to come from the most ancient times and to incorporate the wisdom of the Egyptians, Chaldeans, and Hebrews. When the fashion for reviving ancient thought was at its height, the spurious treatises of "thrice-great

Hermes,” the so-called Hermetic writings, enjoyed great prestige and blended easily with various other secret teachings, such as that of the Jewish Kabbalah.

Toward the end of the Renaissance, the vogue for reviving past philosophies began to subside: Instead, there began to appear “new” philosophies and “new” systems of thought proudly announced as such, for instance, the *Nova de Universis Philosophia* offered by Francesco Patrizzi or the *Great Instauration* (explicitly opposed to a “restoration”) of Francis Bacon. However, most of these efforts at original creation clearly bear the stamp of some ancient sect or sects of philosophy. Even Nicholas of Cusa, the most original systematic mind of the Renaissance, could be called (and indeed once called himself) a Pythagorean. Philosophers hardly ever make a complete break with the past, even when they most loudly claim to be doing so. The great merit of the Renaissance was that thinkers learned what they could from the school of Athens and brought what they learned to bear with fresh vigor upon the problems of human life.

## CARDANO

No individual completely typifies his age, yet it may be useful to focus for a moment on the way in which the various philosophical traditions converged in a single person. As a case history of this sort, we may take the thought of Girolamo Cardano (1501–1576), an Italian medical man and mathematician. Cardano lived in the late, mature stage of the Renaissance, when the dialogues of Plato and the works of Aristotle were known in their entirety, as were Galen and Hippocrates. The Greek commentators on Aristotle were just being recovered and translated. These works were well known at the universities where Cardano studied: Pavia, a stronghold of humanist learning, and Padua, a center of science and medicine. At Padua the biological and logical aspects of Aristotle’s thought were stressed in connection with medical training. Cardano studied under Joannes Montesdoch, a Spaniard, whom he mentions in his writings. There were quite a few such Iberian philosophers studying and teaching in Italy at this time. Aristotelian philosophy was clearly a common European heritage and knew no national boundaries.

A considerable number of Renaissance philosophers were, like Cardano, medical men, and of these quite a few dabbled in mathematics (Galen had urged them to study mathematics for the sake of the training it gave them in sound demonstration). Cardano was, of course, far more successful than most in mathematics: No matter what the true story of his relations with Niccolò Tartaglia may be,

there can be no questioning of Cardano’s grasp of algebra, as shown by his solution of cubic equations. Cardano wrote works on medicine, astrology, and mathematics, but his philosophical reputation must rest primarily on two works in natural philosophy: *De Subtilitate Libri XXI* (On subtlety; 1550) and its sequel, *De Rerum Varietate* (On the variety of things; 1557). *De Subtilitate* attempted a total reconstruction of natural philosophy.

Since other philosophers of the period were inspired to embark on similar projects, it is clear that there was widespread dissatisfaction with Aristotle’s philosophy of nature even before the attacks of Galileo or René Descartes. Aristotle’s physical system was to be threatened dramatically by Copernican heliocentrism, which upset the conceptual scheme on which Aristotle’s analysis of motion was based. This threat was not explicitly posed, however, until the next century, with Galileo’s *Two Chief World Systems*. A Renaissance philosopher such as Cardano did not specifically base his criticisms of Aristotle on the findings of Copernicus or Vesalius: Instead, he reproached Aristotle in a general way for having built up “certain general propositions that experiment teaches to be false.” Cardano presumably intended to remedy this defect, although it must be confessed that his empiricism is not worked out in philosophical detail. This observation would apply with equal force to most Renaissance nature philosophers, few of whom gave more than perfunctory attention to epistemology.

In developing his own system, Cardano started out by taking as his central category something called “subtlety,” which he described as “a certain reason by which sensibilia are with difficulty comprehended by the sense, and intelligibilia by the intellect.” Cardano soon abandons this unpromising concept in favor of a revised Aristotelian terminology in which matter, form, soul, principle, and element play roles somewhat analogous to those they play in Aristotle’s philosophy. For example, Cardano retains the notion of elements but reduces their number from the traditional Aristotelian four to three by eliminating fire, which he classifies as an “accident.” Matter and motion—those central concepts of mechanism—are regarded by Cardano as principles, but they must share this status with form, place, and soul. The last addition puts Cardano into the class of hylozoists, those who believe that all matter is somehow animated, a rather characteristic Renaissance doctrine borrowed largely from Neoplatonism.

Cardano’s writings must have appealed to his Renaissance readers: They are lively, detailed, and full of medical and factual information and misinformation. His style

contrasts sharply with the dry, logically structured argument of the medievals, which can still be found early in the century in the work of a man such as John Major. Cardano obviously delighted in mathematics and in machinery, in this respect, at least, anticipating Galileo in the generation that followed. The amount of superstitious nonsense incorporated in Cardano's work, however, is still distressingly high, and one can easily understand the impatience of later figures such as Gassendi, Thomas Hobbes, and Galileo with their Renaissance predecessors. Cardano wrote a painfully candid autobiography, which appeared in Paris with an evaluation by the French writer Gabriel Naudé (1643). Naudé's judgment on Cardano's character is quite severe. This illustrates a general trend in scholarship: The information current today about many Renaissance thinkers, especially the Italians, comes to us by way of generally hostile French writers of the seventeenth century (Pierre Bayle is exceptional in his lack of polemical intent). If we approach Cardano with the distaste of a Naudé, for example, we too might be inclined to dismiss his work *On Consolation* (1542) as a piece of moralizing cant, when in fact a more humane scholar might consider it a noble document in the light of Cardano's wretched life. Or again, Cardano's passion for gambling could be presented as a despicable and mercenary motive for his interest in games of chance.

But a less censorious approach, such as that of Oystein Ore in his *Cardano, the Gambling Scholar* (Princeton, NJ, 1953), will give Cardano the credit he deserves for anticipating the modern conception of probability as the proportion of favorable outcomes to total possible outcomes. Finally, the mere fact that there was enough interest in Cardano's thought still lingering in seventeenth-century France to justify the publication of his entire work (*Opera Omnia*, 10 vols., Lyons, 1663), shows that Naudé's attitude was by no means universal. This comment could also be made of many other Renaissance philosophers who continued to be read in the seventeenth century, even if not all students of that century were as receptive to Renaissance thought as was Gottfried Wilhelm Leibniz.

**See also** Florentine Academy; Hermeticism; Humanism.

### Bibliography

Everyone interested in the Renaissance should begin by reading two masterpieces of historical writing: Jacob Burckhardt, *Civilization of the Renaissance in Italy* (1st German ed., Basel, 1860), and Johan Huizinga, *The Waning of the Middle Ages* (1st Dutch ed., Haarlem, 1919), both available in various English editions. These works

complement each other: Huizinga deals with France and the Low Countries; Burckhardt deals only with Italy and apologizes for having even mentioned Rabelais. No works of comparable standing in cultural history exist for the Renaissance as it affected England, the German lands, or other European countries.

On the concept of the Renaissance, see Wallace K. Ferguson, *The Renaissance in Historical Thought* (Boston: Houghton Mifflin, 1948), and Franco Simone, *Il rinascimento francese* (Turin: Società Editrice Internazionale, 1961), which adds new evidence for the seventeenth century in France. Invaluable evidence is contained in Herbert Weisinger's articles, especially "The Self-Awareness of the Renaissance," in *Papers of the Michigan Academy of Science, Arts, and Letters* 29 (1944): 561–567. See also Augusto Campana, "The Origin of the Word 'Humanist,'" in *Journal of the Warburg and Courtauld Institutes* 9 (1946): 60–73, and Hans Baron, "The *Querelle* of the Ancients and the Moderns as a Problem for Renaissance Scholarship," in *Journal of the History of Ideas* 20 (1959): 3–22. See also Federico Chabod, "The Concept of the Renaissance," in *Machiavelli and the Renaissance* (New York: Harper Torchbooks, 1965).

For general historical background, Edward M. Hulme, *The Renaissance, the Protestant Revolution, and the Catholic Reformation in Continental Europe* (New York: Century, 1914), is convenient. Excellent summaries may be found in *The New Cambridge Modern History*, Vol. I, *The Renaissance, 1493–1520* (Cambridge, U.K.: Cambridge University Press, 1957).

On humanism, two older works are still basic: Georg Voigt, *Die Wiederbelebung des classischen Alterthums*, 2 vols., 3rd ed. (Berlin, 1893), and J. E. Sandys, *A History of Classical Scholarship*, 3 vols. (Cambridge, U.K.: Cambridge University Press, 1903–1938). All can profit from reading Erwin Panofsky, *Renaissance and Renascences in Western Art* (Stockholm: Almqvist and Wiksell, 1960).

Coming more particularly to philosophy, a useful guide is Paul O. Kristeller and J. H. Randall Jr., "The Study of the Philosophies of the Renaissance," in *Journal of the History of Ideas* 2 (1941): 449–496. Two leading mid-twentieth-century historians of Renaissance philosophy are Paul O. Kristeller and Eugenio Garin. Kristeller's major books are *The Classics and Renaissance Thought* (Cambridge, MA: Oberlin College by Harvard University Press, 1955), revised and enlarged in *Renaissance Thought* (New York, 1961); *Studies in Renaissance Thought and Letters* (Rome: Edizioni di Storia e Letteratura, 1956), which brings together many valuable articles; and *The Philosophy of Marsilio Ficino* (New York: Columbia University Press, 1943). Kristeller's work is solidly based on careful reading of the sources. Garin has dealt with philosophers of the Italian Renaissance extensively in his *La filosofia*, 2 vols. (Milan, 1947), and has compiled useful anthologies, including *Filosofi italiani del quattrocento* (Florence, 1942) and *Il rinascimento italiano* (Milan, 1941). An invaluable anthology, edited by Ernst Cassirer, Kristeller, and Randall, is *The Renaissance Philosophy of Man* (Chicago: University of Chicago Press, 1948).

For French thought during the Renaissance, see Augustin Renaudet, *Préréforme et humanisme à Paris pendant les premières guerres d'Italie (1494–1517)*, 2nd ed. (Paris: Librairie d'Argences, 1953). For Germany, see Peter Petersen, *Geschichte der Aristotelischen Philosophie im*

*protestantischen Deutschland* (Leipzig: Meiner, 1921), and Max Wundt, *Die deutsche Schulmetaphysik des 17. Jahrhunderts* (Tübingen, 1939). On Spain see Carlo Giacon, *La seconda scolastica*, 3 vols. (Milan: Fratelli Bocca, 1944–1950). On Italy, besides Garin's works, see Giuseppe Saitta, *Il pensiero italiano nell'umanesimo e nel rinascimento*, 3 vols., 2nd ed. (Florence: Sansoni, 1961), and Hans Baron, *The Crisis of the Early Italian Renaissance*, 2 vols. (Princeton, NJ: Princeton University Press, 1955). Important studies are those of Ernst Cassirer, especially *The Individual and the Cosmos in Renaissance Philosophy*, translated by Mario Domandi (New York: Harper, 1964).

The Renaissance is a favorite topic for symposia, and the results are sometimes useful: See Tinsley Helton, ed., *The Renaissance* (Madison: University of Wisconsin Press, 1961), with contributions by various authorities. Two specialized studies are Carlo Angelieri, *Il problema religiosa del rinascimento* (Florence, 1952), and Georg Weise, *L'ideale eroico del rinascimento e le sue premesse umanistiche* (Naples: Edizioni Scientifiche Italiane, 1961). Useful articles appear in *Renaissance Studies* and in *Bibliothèque d'humanisme et de renaissance*.

#### OTHER RECOMMENDED WORKS

- Artz, Frederick B. *Renaissance Humanism 1300–1500*. Kent, OH: Kent State University Press, 1966.
- Coates, Willson Havelock, Hayden V. White, and J. Salwyn Schapiro. *The Emergence of Liberal Humanism: An Intellectual History of Western Europe*, Vol. 1. New York: McGraw-Hill, 1966–1970.
- Hankins, James. *Plato in the Italian Renaissance*. Leiden: Brill, 1991.
- Kahn, Victoria. *Rhetoric, Prudence, and Skepticism in the Renaissance*. Ithaca: Cornell University Press, 1985.
- Kelly, Donald R. *Renaissance Humanism*. Boston: Twayne, 1991.
- King, Margaret L. *Humanism, Venice, and Women: Essays on the Italian Renaissance*. Burlington, VT: Ashgate, 2005.
- Kristeller, Paul Oskar, and Mooney, Michael, eds. *Renaissance Thought and Its Sources*. New York: Columbia University Press, 1979.
- Mahoney, Edward P. *Two Aristotelians of the Italian Renaissance: Nicoletto Vernia and Agostino Nifo*. Burlington, VT: Ashgate/Variorum, 2000.
- Moyer, Ann E. *Musica Scientia: Musical Scholarship in the Italian Renaissance*. Ithaca: Cornell University Press, 1992.
- Palisca, Claude V. *Humanism in Italian Renaissance Musical Thought*. New Haven, CT: Yale University Press, 1986.
- Parkinson, G. H. R., ed. *The Renaissance and Seventeenth Century Rationalism*. New York: Routledge, 1993.
- Stinger, Charles L. *Humanism and the Church Fathers: Ambrogio Traversari (1386–1439) and Christian Antiquity in the Italian Renaissance*. Albany: State University of New York Press, 1977.
- Trinkaus, Charles. *In Our Image and Likeness: Humanity and Divinity in Italian Humanist Thought*. Chicago: University of Chicago Press, 1970.

Neal W. Gilbert (1967)

Bibliography updated by Michael J. Farmer (2005)

## RENAN, JOSEPH ERNEST (1823–1892)

Joseph Ernest Renan, the French critic and historian, was born in Tréguier, Brittany. He studied for the priesthood at seminaries in Paris but left the seminary of Saint-Sulpice in 1845 to devote himself to secular teaching and writing. He contributed to the *Revue des deux mondes* from 1851 and the *Journal des débats* from 1853. He received a *docteur ès lettres* in 1852, was elected a member of the Académie des Inscriptions in 1856, and was elected to the Académie Française in 1878. He was appointed professor of Hebrew at the Collège de France in 1862, but the course was then immediately suspended until 1870. In 1884 he became administrator of the Collège de France.

Renan's abandonment of his priestly calling was largely determined by the doubts engendered by his philological study of the Bible. After leaving the seminary, he was strongly influenced by Marcelin Berthelot, the chemist, with whom he maintained a lifelong friendship. Another major influence was German idealism, particularly that of G. W. F. Hegel.

In one sense Renan's life's work can be seen as an attempt to expand the horizons of scientific rationalism by incorporating into it what was valid in idealist philosophy—principally the theme of development, particularly the theme of spontaneous evolution of the human mind. It was the historical aspect and the historical emphasis of Hegel's thought that appealed to Renan, for the cast of his own mind was fundamentally historical, not philosophical. Philosophy for him is not a discipline in its own right, and it is history, not philosophy, that should dominate science; "History is the necessary form of the science of the future." It is evident that Renan used the word *science* in the original sense of "knowledge"; "science" is not to be equated with the natural sciences. On the other hand, his philological and historical method is rationalistic and critical. He was interested, above all, in the evolution of languages and religions as manifestations of the development of the human mind, which is in turn the key to the universe. These manifestations and the universe itself, however, are concrete realities to be discovered through observation, experiment, criticism, and disciplined imagination. They are susceptible to this approach because they are the products of the interplay of natural causes according to constant laws. Renan denied in principle that there is any mystery in the world; what seemed mysterious would yield before the advancing frontiers of knowledge. This is the case in the human no less than in the natural



sciences. Renan, in contact with working scientists, rejected the simplistic notions of natural science characteristic of the positivism of Auguste Comte. He maintained that progress in the natural as well as in the human sciences depends on human judgments of the balance of probabilities on the evidence. He further maintained that all reality is in some degree historical, that the natural sciences (paleontology, for example) reveal the remote parts of history, and that the human and natural sciences can and must therefore be of mutual help.

Just as he banished all traditional metaphysics from philosophy, Renan rejected any supernatural content in religion. The true religion of humankind, in the sense of "a belief accompanied by enthusiasm which crowns conviction with devotion and faith with sacrifice," is that of science (that is, knowledge). Renan's argument runs as follows: The universe is characterized by change according to "laws of progress" under which the human mind becomes increasingly conscious of itself and the ideal is increasingly manifested amid the real: "The goal of the world is the development of mind." At the end of the process God, in the sense not of a creative providence but of an immanent ideal, will be realized. Since this ideal consists in the complete development of consciousness and in the attainment by that consciousness of the full measure of beauty and morality of which it is capable, science must be the great task of humankind. This task must be approached in the spirit neither of mere curiosity nor of mere utilitarianism but in the true religious spirit, seeking revelation of the divine.

The above sketch of Renan's thought is based mainly on his youthful work, *L'avenir de la science*, written in 1848 but first published in 1890. In his later philosophical writing he modified, but did not abandon, the fundamental position adopted there. Political and social events in France, in particular, damped his optimism and strengthened his skeptical and ironical streak. He began to have doubts about the "religion of science" to which he had turned when he abandoned Roman Catholicism. He became less sure that men had the capacity to attain adequate knowledge, and some of his own writing became tentative, cast at times in the form of dialogue. In his professional historical work, however, which always remained his chief concern, he stood fast by his views on the development of rationality out of instinct and on the progressive realization of God on Earth. Even in the new preface that he added to *L'avenir de la science* on its publication late in his life, Renan declared that his religion was still "the progress of reason, that is to say, of science." He had been too sanguine, too anthropocentric, and not

entirely emancipated from Catholicism; the growth of knowledge had not, in fact, clarified human destiny. He confessed that he did not see how humankind could maintain its ideals if deprived of its illusions, but he retained his faith in knowledge as the supreme pursuit.

**See also** Hegel, Georg Wilhelm Friedrich; Idealism; Modernism; Natural Law; Rationalism.

### Bibliography

#### WORKS BY RENAN

- Averroes et l'averroïsme*. Paris: Durand, 1852.  
*Essais de morale et de critique*. Paris: Calmann-Lévy, 1859.  
*Vie de Jésus*. Paris: Michel Lévy, 1863. Translated by C. E. Wilbour as *The Life of Jesus*. New York: Carleton, 1864.  
*Questions contemporaines*. Paris: n.p., 1868.  
*Dialogues et fragments philosophiques*. Paris: Calmann-Lévy, 1876.  
*Souvenirs d'enfance et de jeunesse*. Paris: Calmann-Lévy, 1883. Translated by C. B. Pitman as *Recollections of My Youth*. Boston and New York: Houghton Mifflin, 1929.  
*L'avenir de la science*. Paris: Calmann-Lévy, 1890. Translated by A. Vandam and C. Pitman as *The Future of Science*. Boston: Roberts Brothers, 1891.

#### WORKS ON RENAN

- Allier, Raoul. *La philosophie d'Ernest Renan*. Paris, 1895.  
 Berthelot, René. "La pensée philosophique de Renan." *Revue de métaphysique et de morale* 30 (1923): 365–388.  
 Charlton, D. G. *Positivist Thought in France during the Second Empire*. Oxford: Clarendon Press, 1959. See Ch. 6.  
 Cresson, André. *Ernest Renan: Sa Vie, son oeuvre, avec un exposé de sa philosophie*. Paris: Presses Universitaires de France, 1949.  
 Mott, Lewis Freeman. *Ernest Renan*. New York and London: Appleton, 1921.  
 Séailles, Gabriel. *Ernest Renan: Essai de biographie psychologique*. Paris: Perrin, 1895.

W. M. Simon (1967)

## RENOUVIER, CHARLES BERNARD (1815–1903)

The French critical philosopher Charles Bernard Renouvier was born in Montpellier and was educated at the École Polytechnique, where he specialized in mathematics and natural science. At the school he came under the influence of the work of Antoine Cournot and of Auguste Comte, who at that time was an instructor (*répétiteur*) in higher mathematics there. In 1848 Renouvier published in Paris his *Manuel républicain de l'homme et du citoyen*, a volume addressed to schoolteachers, which urged the

preaching of socialism. But his political views were frustrated by the coup d'état of Louis Napoleon, and he retired from active participation in politics to write philosophy. Renouvier never held an academic position but worked as a private individual, producing one of the longest series of philosophical works in French history. In 1867 he began the publication, with his friend and collaborator François Pillon, of *L'année philosophique*, a monthly that propagated Renouvier's philosophical doctrines. These doctrines were chiefly expounded in a series of books, constantly revised by Renouvier, the *Essais de critique générale*, the final edition of which appeared in 1897. He continued writing up to the time of his death, his last work being *Le personalisme* (1903). Though his pluralism and his personalism anticipated some philosophical doctrines of the early twentieth century, his main influence was upon his French contemporaries.

### NEOCRITISM

Renouvier's general position is called neocriticism, because it took the method of Immanuel Kant's critical philosophy as its starting point. But though Renouvier started with Kant's method, he did not accept Kant's conclusions but used them rather as a basis from which to launch a set of ideas often critical of Kant.

Renouvier laid down as an integral part of his philosophy what he called the "law of numbers," according to which every cardinal number is an ultimate individual, finite and irreducible. Mathematics is the paradigm of thinking, and the law of contradiction is more clearly manifested in mathematical operations than anywhere else. But the term *mathematics*, as Renouvier used it, was restricted to arithmetic, and he derived the nature of numbers exclusively from the cardinal numbers. This led him to deny the existence of any infinite, for he maintained—unable to anticipate the work of Georg Cantor—that an infinite number was a contradiction in terms. Renouvier extended his criticism of the notion of infinity beyond numbers to deny the infinity of space and time as well.

Renouvier recognized that knowledge is relative to its premises and to the person who laid down the premises; nevertheless he could not accept the relativity of logical processes. There is a distinction involved here between logic and the psychology of thought. Just as each number is a distinct and separate entity, so is each human being. And just as the characteristics of each number—duality, triplicity, and so on—can never be reduced to, or "reconciled" with, the characteristics of any other number, so each human being is not exactly like any other and can-

not be merged into a general group-consciousness or absorbed into an absolute mind. Knowledge is always the property of individual knowers, and the distinction between knowledge and belief disappears. What an individual knows is what seems reasonable to him, and his contribution to knowledge can never be subtracted. The subtraction can be made verbally, to be sure, but to do so is to alter the character of cognition, which is essentially judgment.

### PHENOMENA

Renouvier also differed from Kant in his doctrine of phenomena. Phenomena are not the appearances of anything other than themselves. They are neither illusions nor purely subjective beings. They are *sui generis*, being whatever we perceive or whatever we make judgments about. He granted that the name is unfortunate except insofar as it indicates appearances. Because there are no things-in-themselves, Renouvier criticized Kant's antinomies, which hold good only if there are noumena. His attack on the first antinomy, for example, was based on its use of the concept of infinity. Since infinity is an inherently inconsistent idea, Renouvier asserted that the world must have had a beginning in time and that space is limited. The domination of the number concept as a conceptual model appears here in full force. For Renouvier, the numbers begin with one, since zero and negative numbers are not really numbers, and spaces are the spaces of individual discrete beings, there being no such entity as number-in-itself or space-in-itself.

There exists within the number series the category of relation. For the numbers are ordered, and order is a kind of relation. All other categories are, for Renouvier, forms of relation, but of relation as discovered within the framework of an individual's consciousness. There turn out to be nine categories—relation, number, position, succession, quality, becoming (*devenir*), causality, purposiveness, and personality. Each has its thesis, antithesis, and synthesis; and all are rooted in the phenomenal world as judged by us. It is uncertain whether Renouvier attempted to derive his categories in the manner of Maine de Biran from personality—our acting as a cause, our seeking ends, our sensory discriminations (which might produce the separateness of quantity and quality), spatial positions, moments in time, and the intervals between them—or whether his assertion that personality is one of the categories is derived from his premise of the law of numbers. In any event, just as each number has its own distinctive quality, its own position in the numeral order, and its many relations to other numbers, determined not

only by its own character but also by that of the other numbers, so the human being has his own personality and displays the other categories not only as a distinct entity but also as a perceptive consciousness.

The parallelism between the ways in which a man judges, perceives, and knows and the ways in which he as a person differs from other beings pervades Renouvier's writings. Thus, because one acts to achieve one's purposes, it follows that both causality and purposiveness exist within the human being and must likewise be combined in the phenomenal world. A cause determines the path of an event, but the direction of that event is determined by that which participates in it.

Since no two events are exactly alike, the deterministic factor in nature is mitigated by chance. Renouvier probably got this argument from Cournot, who also insisted upon the probabilistic element in nature. To frame a law or a generalized description depends upon our ability to discover absolutely homogeneous phenomena or groups of phenomena. If this is impossible, then generalizations are at most only probable. But at the same time, each individual phenomenon contributes something to the events of which it is a part, and that contribution in the very nature of things cannot be predicted.

## INDETERMINISM

The problem of causation arises with regard to human beings in the form of the antithesis between free will and determinism. Since every act of consciousness is a relation between a perceiving subject and that which is perceptible, then as soon as a conscious act is formulated and made clear to the perceiving mind, it will be organized in terms of the categories. But there is a choice among the various categories to be applied, for we are not forced either to quantify or qualify, to count or to locate, to assign a date or to recognize a cause. The categories limit the possibilities of judgment but have no inherent order of predominance. In other words, Renouvier held that when we see a phenomenon, for example, a tree, we are not forced first to judge it as green, then as distant, old, fan-shaped, simple, or what you will. The order of judgment is determined by us, and we are free, within the range of possible categories, to judge it as we will. The selection of a category or group of categories depends on our free choice in accordance with our interests at the moment of judging.

Freedom cannot be proved, nor can determinism. Both are assumptions utilized in view of their consequences. These consequences may be purely intellectual or may be moral or practical. But freedom itself rests

upon the inherent individuality of the human will, an individuality which cannot be completely absorbed into any larger class of beings. Insofar as any being is unique, to that extent it is undetermined or self-determined. And insofar as it is identical with other beings, to that extent the homogeneity of its class accounts for the regularity of its behavior. In short, individuality and freedom are synonymous terms, and Renouvier even called freedom the principle of individuation. The consequence is that just as the personal equation enters into all judgments, so the only certainty we have is the certainty of our judgments. Renouvier put it as follows:

Certitude is not and cannot be absolute. It is a condition and act of man—not an act and a condition in which he grasps immediately that which could not be immediate, i.e., facts and laws external and superior to present experience, but rather one in which he posits his awareness as it exists and as he maintains it. Strictly speaking, there is no certitude; there are only men who are certain. (*Traité de psychologie rationnelle*, Paris, 1912, Vol. I, p. 366)

But indeterminism is not limited to human judgments. It extends also to history. For since history is in part made up of human behavior, human decisions must be included in its scope, and there is no way of eliminating them. One can, of course, describe the environment of human life, its stability, and its mutability; but if it remains stable, that is because human beings have not changed it, and if it changes, that is due to human acts as much as to natural disasters. People modify their living conditions, not as a group acting as one person, but as a collection of individuals. Their reasons for doing so may vary, as is inevitable, and of course they are not able to modify their conditions completely. But Renouvier emphasized the importance of human decisions for the way in which individuals will live, since the ability of human beings to make choices makes it impossible to lay down either a law of universal progress toward the good or one of constant degeneration. Hence Renouvier rejected historical laws, such as those of Comte and G. W. F. Hegel, though he was attracted to meliorism.

## ETHICS

If there is no historical law dooming humankind to move in any predetermined direction and if history only records actual change, the question arises of the relation of history to ethics. People make moral judgments and act so as to achieve what they believe to be right. Morals, then, are not the result of history, though what happens

in history reflects our moral judgments. Morals are rather the source of historical changes, and if we are to appraise historical events, we shall have to do so in moral terms. This clearly requires a definition of good and evil, and in view of the radical individualism of Renouvier this might seem an insurmountable task. But he identified evil with conflict, conflict both between persons and between groups of persons. For warfare is in essence the prevention by one or more persons of the fulfillment of the volitions of others. Hence tyranny, slavery, and conquest are to be condemned. This assumes that it is possible for a group of enlightened people to respect the individuality of their fellows and for all to live in peace.

In his fictional account of what history might have been, *Uchronie* (1876), Renouvier claimed that the secret of human happiness lies in our recognition of the individual's freedom. If at any epoch people had accepted individual freedom wholeheartedly, he argued, universal peace and harmony would have prevailed. Religious, economic, and national wars would have ended at once; for everyone would have taken it for granted that each person has a right to his own religious views, to the satisfaction of his own economic interests, and to his own national loyalties. Renouvier held that education alone could bring this about, though he had no illusions that proper education was ever likely to be instituted. The dogma of historical determinism has had too firm a hold on human will power and has brought about acquiescence, sloth, injustice, and ignorance.

The basic premises of Renouvier's *Science de la morale* (1869) are that human nature is rational and that people believe themselves to be free. Their belief in freedom leads individuals to act for what they judge to be better, and their rationality guides them in their choice of ends. To act morally is to act rationally. By doing so we rise above the beasts; we recognize the humanity in our fellows and respect it. For this reason Renouvier became a bitter opponent of the Catholic Church and of monarchy and urged his readers to turn to Protestantism as the religion of individual conscience. To him Protestantism was the religion of a personal God—not an absolute and unchanging Being, omniscient and omnipotent, but finite, limited, free, and the guarantor of our freedom. God's existence is not proved, but it is a reasonable hypothesis drawn from the existence of our moral objectives. Running through Renouvier's many works are the premises that the plurality of existing things is irreducible; that chance is real and is reproduced in individual freedom of choice; that time and novelty really exist; and that no absolutes or infinities exist.

*See also* Comte, Auguste; Cournot, Antoine Augustin; Determinism and Indeterminism; Hegel, Georg Wilhelm Friedrich; History and Historiography of Philosophy; Infinity in Mathematics and Logic; Kant, Immanuel; Maine de Biran; Neo-Kantianism.

## Bibliography

### WORKS BY RENOUVIER

- Essais de critique générale*. 4 vols. Paris, 1854–1864.  
*Science de la morale*. 2 vols. Paris, 1869.  
*Uchronie, l'utopie dans l'histoire*. Paris, 1876.  
*Esquisse d'une classification systématique des doctrines philosophiques*. 2 vols. Paris, 1885–1886.  
*Philosophique analytique de l'histoire*. Paris, 1896–1897.  
*Le personalisme*. Paris, 1903.  
*Lettres de Renouvier à Hennequy (1861–1899)*, edited by Marcel Méry. Gap: Editions Ophrys, 1969.  
*Manuel républicain de l'homme et du citoyen* (1848). Paris: Garnier, 1981.  
*Les dilemmes de la métaphysique pure*. 2nd ed. Paris: Presses universitaires de France, 1991.  
*Victor Hugo, le philosophe*. Paris: Maisonneuve et Larose, 2002.

### WORKS ON RENOUVIER

- Blais, Marie-Claude. *Au principe de la République: Le Cas Renouvier*. Paris: Gallimard, 2000.  
 Fedi, Laurent. *Le problème de la connaissance dans la philosophie de Charles Renouvier*. Paris: L'Harmattan, 1998.  
 Foucher, Louis. *La jeunesse de Renouvier et sa première philosophie*. Paris: Vrin, 1927.  
 Grande, Sonia Kase. *Charles Renouvier and the Search for Intellectual and Social Synthesis in Nineteenth-Century France*. 1975.  
 Gunn, J. Alexander. *Modern French Philosophy*. London: Unwin, 1922. Contains an excellent bibliography.  
 Hamelin, Octave. *Le système de Renouvier*. Paris: Vrin, 1927.  
 Logue, William. *Charles Renouvier, Philosopher of Liberty*. Baton Rouge: Louisiana State University Press, 1993.  
 Milhaud, Gaston. *La philosophie de Charles Renouvier*. Paris, 1927.  
 Séailles, Gabriel. *La philosophie de Charles Renouvier*. Paris: Alcan, 1905.  
 Turlot, Fernand. *Le personalisme critique de Charles Renouvier: Une philosophie française*. Strasbourg: Presses Universitaires de Strasbourg, 2003.

George Boas (1967)

Bibliography updated by Michael J. Farmer (2005)

## RENSI, GIUSEPPE (1870–1941)

Giuseppe Rensi was an Italian skeptical philosopher and professor of philosophy at the universities of Messina and

Genoa. Rensi first upheld a religiously or theistically oriented idealistic philosophy, defending it in a number of essays and fostering it through his translations of the works of Josiah Royce. He contrasted his theistic “constructive idealism” with the “immanentistic idealism” of Benedetto Croce and Giovanni Gentile; he regarded the latter as a temporary position that, if developed coherently, would have led to constructive idealism. According to Rensi, an idealism that does not arrive at God subtracts reality both from the external world, which then becomes a set of ideas, and from the human spirit, which is then resolved into a set of ideas without a subject.

After World War I, regarded by Rensi as proof of the fundamental irrationality of the world, he began to defend a radical skepticism based on the multiplicity, irreducibility, and irreconcilability of opinions, the reasons used to justify them, and some aesthetic tastes and moral ideas. Rensi held that the traditional objection to skepticism—that it contradicts itself by asserting that there is no truth while dogmatically asserting its own truth—was a purely verbal objection, because the skeptic holds his position against any doctrine taken in itself by showing the contradictions and shortcomings of that doctrine. Therefore the skeptic does not assert that there is no truth but instead that a particular doctrine that claims to possess truth does not and cannot possess truth. Skepticism, in other words, shows the disagreement of reason with itself both within the views of one man and between the views of different individuals. War, the conflict of rights and of political powers, and the contradictory character of philosophies are, according to Rensi, proofs of the intrinsic contradiction in reason. Skepticism does not exclude faith but stems from the preservation of faith. The skeptic is skeptical not because he does not believe but because others believe differently than he; that is, they believe that which he considers absurd.

Rensi had been a socialist in his youth but later came to defend authority. He wished to give to power (and even violence) the function of helping man escape from the chaos of opinions and contrasting interests and of forming a people into an economic, political, and spiritual unity. Authority need not base itself on reason, because it creates for itself the reason of all that it wishes. Although these ideas seem close to those of fascism, Rensi quickly declared himself opposed to fascism and remained so until his death.

According to Rensi, skepticism implies atheism in the field of religion. The refinement of religion that leads to regarding God as inaccessible to the senses and to human powers makes God a nonbeing, the pure and sim-

ple negation of every reality accessible to man. From this point of view, both negative theology and mysticism demonstrate atheism. Atheism is still a religion because it is an answer—even if a negative one—to the problem of supreme reality. Unlike other religions, atheism is absolutely disinterested because it contains no egoistic motive and because it places man before the mystery of the All without his being able to expect from the All any help for his own needs.

After 1922, when the absolute idealism of Croce and Gentile assumed the status of an official or semiofficial philosophy in Italy, Rensi accentuated his polemic against it and affirmed the theses most opposed to those of idealism: materialism and pessimism. The Kantian system, considered to be idealistic by the idealists, seemed to Rensi to justify materialism because the Kantian forms of intuition and of thought that condition phenomena, and therefore the totality of nature, are not created by the self but constitute “consciousness in general,” which is the intelligibility of the things themselves. According to Rensi, the Kantian doctrine is, therefore, that nature gives reality and knowability to natural things, that things generate of themselves, and of themselves are spatial, temporal, perceptible, and representable; in one word, they are material.

Rensi held that materialism implies pessimism because a material nature deprived of any finality offers man no guarantee and necessarily includes evil, error, and conflicts. For a man who lives in such a nature, morality, when not based on an egoistic calculus or subjected to an imposed code, is a disinterested recognition of evil and a protest against it. It is therefore pure folly. Nevertheless, all of Rensi’s works contain a mystical and religious strain, a sense of mystery and of a force that, the triumph of evil in nature and in history notwithstanding, reveals itself in the interiority of man. Rensi condensed this feeling into the phrase “Atoms and the void—and the divine in me.”

**See also** Atheism; Croce, Benedetto; Gentile, Giovanni; Idealism; Materialism; Pessimism and Optimism; Royce, Josiah; Skepticism.

### Bibliography

#### WORKS BY RENSI

*Il genio etico e altri saggi.* Bari, 1912.

*La trascendenza.* Turin, 1914.

*Lineamenti di filosofia scettica.* Bologna, 1919.

*La filosofia dell' autorità.* Palermo, 1920.

*La scepsi estetica.* Bologna: N. Zanichelli, 1920.

*Introduzione alla scepsi etica.* Florence, 1921.  
*Apologia dell'ateismo.* Rome, 1925.  
*Realismo.* Rome, 1925.  
*Il materialismo critico.* Rome, 1927.  
*La morale come pazzia.* Modena: Guanda, 1942.

## WORKS ON RENSI

Morra, G. *Scetticismo e misticismo nel pensiero di Giuseppe Rensi.* Palermo, 1958.  
 Nonis, P. *La scepsi etica, di Giuseppe Rensi.* Rome: Studium, 1957.

**Nicola Abbagnano (1967)**  
*Translated by Nicholas Spadaccini*

## REPRESENTATIVE REALISM

See *Realism*

## REPUBLICANISM

Republicanism is one of the great traditions of Western political thought. To say that republicanism is a “tradition” of political thought is to say that distinctively republican ideas about politics have been championed by a number of authors in the history of political theorizing, and that many of the later authors who championed those ideas consciously drew on and developed the work of earlier ones. This continuity of reference and influence makes it possible to trace a republican strand in Western political writing. But what ideas about politics are distinctively republican? What ideas define the republican tradition?

The republican tradition is often associated with the claims that citizens can only be free in a free society, that the opposite of freedom is a state of dependence akin to slavery, that societies are most likely to enjoy freedom and to realize their common good when they are governed by politically engaged citizens who act from the civic virtues, and that the pursuit of the common good is undermined when citizens’ virtues are corrupted by selfishness, luxury, and ambition. These claims turn up consistently in the writings which make up the republican tradition from Rome at least through the eighteenth century.

The classic texts of the republican tradition were produced in political circumstances very different than those of the early twenty-first century. These texts commend ways of life, such as a life of politically active citizenship, that are open to relatively few citizens of large, modern societies such as England and the United States. The political threats in the face of which these texts were

produced were quite different than the threats to liberty and equality posed by the modern states of late capitalism. Republicans’ emphasis on civic virtues raises the possibility that republican politics would be difficult to sustain under conditions of moral pluralism. It is therefore not immediately clear that republicanism can provide guidance to modern politics. Even if republican ideas can provide some guidance, it is far from clear that republicanism alone can provide sufficient guidance. Perhaps republican ideas about politics are most useful as supplements to political theories that belong to other traditions of political thought and that are explicitly framed for current conditions.

Republicanism has enjoyed a revival in legal and political philosophy since the 1980s. Those who have revived republicanism in these disciplines have tried to apply the insights and arguments of the tradition to contemporary politics. But some participants in the revival themselves seem to raise questions about the sufficiency and distinctiveness of republicanism. They move easily between republicanism and democratic liberalism and seem content to describe themselves as both republicans and liberals. The fact that they do so raises questions about whether any version of republicanism that is of more than historical interest is faithful to ideas that have distinguished the republican tradition. It also raises questions about whether versions of republicanism that bear on contemporary politics are part of a strand that ought to remain distinct from other movements of political thought. Perhaps the insights of republicanism are best absorbed into liberalism, a tradition which has its origins in the early modern period.

Late twentieth-century work on republican views of liberty has, however, changed the way historians and political theorists think about republicanism. It has already shed new light on some of the classic texts in the republican tradition. It promises to show how some of the claims characteristic of republican writing can be systematically united and given a theoretical basis. And it promises to illuminate deep and interesting differences between republicanism and liberalism. If these promises can be made good, then republicanism’s claim to be a distinctive family of political thought—and one of continuing relevance—can be vindicated.

## ROMAN REPUBLICANISM

The origins of the republican tradition lie in the writings of Roman political thinkers, such as Cicero and Sallust, who lamented and analyzed Rome’s transformation into an empire just before the beginning of the common era.

They came to be called “republicans” in part because the form of government they favored was that of pre-imperial Rome—a regime for which Cicero popularized the name “the republic.” They are also called “republicans” because of the features of that government that they seized upon when arguing that rule by the republic’s government was superior to imperial rule. The words “republic” and “republicanism” derive from the Latin phrase *res publica*, which means “public matter.”

According to these thinkers, the republic was better suited to advance the common good of the Roman people than the empire was because, unlike the empire, its government was participatory. It was governed by public-spirited citizens—in particular, public-spirited citizens serving in the Roman senate—who devoted themselves to the pursuit of public matters rather than to the pursuit of their own wealth and ambition. In the “Dream of Scipio,” Cicero famously claimed that those who dedicate themselves to the preservation of the republic would enjoy an eternal reward. Such devotion to the republic, he thought, required civic virtue. Republican writers claimed that the Roman republic was subverted by corruption. It was subverted, they said, by those who sought and used political power to further their own ends rather than the common good of the Roman people.

## RENAISSANCE AND EARLY MODERN REPUBLICANISM

Republicanism had little impact on the political thought of the Middle Ages, though some of the writings of Cicero were certainly known to such great medieval philosophers as Augustine and Thomas Aquinas. But the writings of the Roman republicans were important sources for political thinkers in Renaissance Italy who wanted to maintain the freedom of city-states against internal and external threats. They were also important sources for thinkers in seventeenth-century England who opposed the absolutist tendencies of the Crown. These writers located themselves in the tradition of Roman republicanism. They drew on republican claims about the importance of political participation, the need for a virtuous citizenry and the threats posed by corruption and self-interest, even as they adapted those claims to their own situations.

Among Italian thinkers, the greatest was undoubtedly Machiavelli, especially the Machiavelli who wrote *The Discourses*. Machiavelli believed that the citizens could only enjoy freedom if their city was free. One of the most significant threats to a city’s freedom, he argued, was an internal threat: the threat posed to a city’s good

government by factions that would pursue their own interests once in power. Inspired by the political ideas of the humanist tradition and the writings of Roman republicans, Machiavelli argued that the dangers of factionalism could best be averted by a government of citizens committed to the common goods of civic wealth, glory, and independence.

English republicans such as John Milton and James Harrington were less concerned with the threat of faction than they were with what they regarded as the absolutist tendencies of the monarchy. They maintained that absolute power corrupted, but it did not corrupt only the monarch. A powerful court, they thought, was one that corrupted courtiers and politicians by encouraging their dependence upon royal favor. English republicans stressed the importance of the civic virtues, among which they numbered independence and frugality. They held up as models of good government the republics of Renaissance Italy, and the Roman republic. Historians sometimes call them “commonwealth men” because of their support for the Puritan commonwealth.

## THE REPUBLICAN REVIVAL

The last third of the twentieth century saw a resurgence of scholarly interest in republicanism, primarily but not exclusively in the English-speaking world. The resurgence of interest among American Constitutional lawyers in the 1980s and 1990s came to be known as the “republican revival.” That term can be stretched to encompass the contemporaneous revival of interest in republicanism among political philosophers. The republican revival among lawyers and philosophers was preceded by and drew upon work by historians of Renaissance political thought and by historians of the American founding. Indeed it was because of the resurgence of interest among historians that so much has been learned about early modern republicanism. It is useful to begin a survey of the republican revival with a look at some of the historical work that preceded and influenced legal philosophers responsible for the revival.

In the 1950s Louis Hartz articulated what was for a time the received orthodoxy about the intellectual foundations of the American Revolution and founding. According to Hartz, the revolutionaries and founders owed their greatest intellectual debts to the classical liberalism Hartz ascribed to John Locke. In the 1960s historians of the American founding and its intellectual antecedents, notably Bernard Bailyn and Gordon Wood, raised serious challenges to this orthodoxy. Bailyn and Wood argued powerfully that the intellectual underpin-

nings of the revolution and the founding period were in large part republican, drawn from the English commonwealth tradition of the previous century. John Pocock, who traced the origins of the commonwealth thought to Renaissance Italy, provided an even longer genealogy for American republicanism.

Bailyn and Wood mined the pamphlets and popular literature of early America for evidence of republican political thinking. The expressions of republicanism they found there included pervasive emphasis on the need for citizens to dedicate themselves to the common good and on the deleterious effect of faction and the elevation of private over public interest, concern with the corrupting effects of various forms of dependence upon Britain (including dependence on its monied and manufacturing interests), and the description of American dependence as a condition of slavery. Pocock, Bailyn, and Wood all maintained that the republicanism of the American founding was only gradually eclipsed by other forms of political thought in the years or decades that followed.

The question of whether and to what extent the American founders were republicans is a question of some importance for legal philosophy. The founding period of the United States was the period in which the body of the American Constitution and the Bill of Rights were written. The conclusion that the founders owed deep intellectual debts to republicanism arguably has profound implications for how the Constitution and the Bill of Rights should be read and applied. The argument that it has such implications seemed especially pressing to legal scholars at a time when some were defending originalist canons of Constitutional interpretation. In the 1980s Constitutional scholars began to draw on the historical work of Pocock, Bailyn, Wood, and others, and initiated the republican revival in legal scholarship.

The leading figures of this revival, such as Cass Sunstein and Frank Michelman, emphasized the participatory strain of republicanism. Republican government, according to these thinkers, is government by citizens who participate in politics. The politics in which they are to participate is to be deliberative: citizens of a republican regime are to participate in collective deliberations about public matters. Such public deliberation, they argued, promises to combat the factionalism and self-interest that republicans had traditionally seen as undermining good government. It does so because the process of deliberating with others is not one of bargaining in which parties try to satisfy the preferences they have formed before public deliberation begins. Rather, it is to be a process of reasoning with others about how to advance the common

good. When citizens reason together about the common good, they are forced to rethink whatever self- and group-interested preferences they may bring into public deliberation.

Republican accounts of politics had previously been addressed to societies much smaller than the democracies of the late twentieth century. Framing a version of republicanism adequate for such large societies required imagining institutional forms through which republican government could be exercised within them. The leaders of the republican revival in the law offered republican readings of the American constitution and drew out the implications of those readings for a host of questions in public law, from environmental law to campaign finance reform.

The republicanism offered by republican revivalists in the legal academy—like the republicanism uncovered by historians of the American founding—emphasized the value of political participation, the importance of their commitment to the common good and the threat posed by citizens' unregulated pursuit of self- and group-interested preferences. Because of these emphases, republicanism seemed to offer a healthy corrective to the individualism, self-interest, acquisitiveness, and withdrawal from public life that some thinkers, such as Michael Sandel, have alleged that liberalism encourages. Yet the republicans in the legal academy saw significant continuities between their own views and some forms of liberalism, particularly between their own views and the version of liberalism developed and refined by John Rawls from the 1960s until his death in 2003. By a decade after the republican revival began in American law schools, some of its leading figures had ceased to insist that there was anything distinctively republican about their views. Even some who continued to describe themselves as republicans, such as Sunstein, also described themselves as liberals and “deliberative democrats” as well. The development of a republicanism that was explicitly contrasted with liberalism had to await the republican revival in constructive political philosophy.

According to some leading republican political philosophers, the differences between republicanism and liberalism lie, not in the former's emphasis on political participation and civic virtue or in the latter's emphasis on individual rights, but in the very different conceptions of political freedom associated with each. Liberalism, as its name suggests, is a political philosophy that values liberty. The liberty that liberals are sometimes said to value is what has come to be called “negative liberty.” Someone enjoys negative liberty to the extent that she can act as she



likes, without external impediments. Political liberty is the freedom citizens enjoy in political society. Those who identify political liberty with negative liberty must think that even the best law is an external impediment to action, and so interferes with citizens' political liberty. If they also think, as liberals do, that it is the job of government to promote and secure political liberty, then they must also think government should rely as little on these impediments as possible. It should secure as much negative liberty for citizens as is compatible with the enforcement of laws needed to maintain public order.

Some of the most prominent republicans who have contrasted their views with liberalism have contrasted them with versions of liberalism which equate political liberty with negative liberty. They have introduced another kind of freedom, which they call "freedom as nondomination." They have argued either that political liberty includes both negative liberty and liberty as nondomination, or that it consists in liberty as nondomination alone. To appreciate the differences these republicans see between liberalism and their own views, it is necessary to see what it is for one agent to dominate another.

One agent dominates another just in case the former is in a position to interfere arbitrarily with the choices of the latter. An agent is in a position to interfere arbitrarily with another's choices just in case that agent is able to interfere with the other's choices without having to take the latter's interests into account. This way of characterizing domination implies that there are two important differences between liberal views which identify political liberty with negative liberty and republican views which either equate political liberty with liberty as nondomination, or which claim that political liberty includes liberty as nondomination.

One difference is that, according to the latter, not all laws restrict citizens' freedom. When political authorities take account of the interests of citizens in the enactment and enforcement of law, they do not dominate citizens. They do not dominate them because, though the laws may interfere with citizens' freedom of action, they do not do so arbitrarily. Therefore, though these authorities compromise citizens' freedom on liberal accounts which equate liberty with negative liberty, they do not do so on republican accounts.

The other difference is that republicans think one person can restrict another's liberty just by being in a position to interfere with him arbitrarily, even if she never actually interferes with him at all. Thus a political authority who can exercise power without accountability, but who chooses to enact laws which further the common

good, still dominates citizens. These citizens therefore lack political freedom on republican accounts but not on liberal ones.

The account of liberty as nondomination has been stated and defended most notably by Philip Pettit, beginning in the mid-1990s. Pettit labels that account of freedom a "republican" account because he claims that it is found in the seminal texts of the republican tradition. He argues quite convincingly that, by taking freedom understood as nondomination as the supreme political value, he can account for why republicans have valued political participation and why they have maintained that citizens are free only in free societies. Quentin Skinner, the historian of Renaissance republicanism, also has claimed to have found a distinctive conception of liberty in the republican tradition. In response to Pettit's work, Skinner has argued that republican political liberty includes both negative liberty and liberty as nondomination. Whether Pettit's or Skinner's view of political liberty is more faithful to the texts remains a matter of scholarly debate. What is beyond debate is that Pettit's conceptual work on republican liberty has greatly influenced historical work on republicanism, and that Pettit and Skinner have taken the republican revival to a new level of philosophical sophistication. Questions remain, however, about exactly where their versions of republicanism differ from prominent forms of liberalism which do not equate political liberty with negative liberty.

The liberalism developed by Rawls in the last third of the twentieth century has been enormously influential. Rawls argued for principles of justice which, he maintained, would be agreed to in what he called "the original position." The original position is, like the state of nature in Locke's work, a condition appropriate for writing a social contract. Thus the principles of justice Rawls defends are principles citizens would choose for themselves under the conditions appropriate for such a choice. Rawls calls a society which is regulated by those principles a "well-ordered society." When citizens live in a well-ordered society and when their own plans are in accord with the principles of justice, they live under and act from principles they would give to themselves. To be autonomous is, literally, to give oneself a law. Citizens who live in a well-ordered society and act from principles they would give themselves therefore enjoy an important form of autonomy, which Rawls calls "political autonomy."

One question republicans need to answer is how their view of political liberty differs from a view of political liberty like Rawls's, according to which political lib-

erty includes political autonomy as an important ingredient. Another is whether they think Rawls's well-ordered society would include injustices or obstacles to freedom—for example, instances of domination—that a republican regime would not. This question will require a complicated answer because, while Rawls thinks that political freedom includes political autonomy, he does not equate the two. He insists that, in a well-ordered society, the liberties exercised in political participation will have what he calls “fair value.” Citizens' possession of political liberties which have fair value may be required by their political autonomy, but it seems to be distinct form of political freedom. Moreover, it is at least arguable that when citizens enjoy this form of political freedom—when these liberties have fair value—much of the political domination that concerns republicans will be eliminated. Finally, republicans need to ask whether a republican society would allow injustices that a well-ordered society would eliminate.

Republicanism has undoubtedly been a philosophically interesting tradition of thought which exercised great influence at important points in Western political history. Since the late twentieth century it has been revived with brilliance and ingenuity. But until contemporary republicans answer these questions, it will be difficult for them to maintain that republicanism is superior to all forms of liberalism. Until they answer them, it will also be unclear whether republicanism can stand on its own as a theory of contemporary interest, or whether the insights that have been systematized by its most sophisticated exponents are better incorporated into some version of liberal theory.

**See also** Cosmopolitanism; Libertarianism; Multiculturalism.

### **Bibliography**

- Bailyn, Bernard. *The Ideological Origins of the American Revolution*. Cambridge, MA: Harvard University Press, 1992.
- Cicero, Marcus Tullius. *De re Publica*. Translated by Clinton Walker Keyes. Cambridge, MA: Harvard University Press, 1959.
- Harrington, James. “*The Commonwealth of Oceana*” and “*A System of Politics*.” Cambridge, U.K.: Cambridge University Press, 1992.
- Hartz, Louis. *The Liberal Tradition in America*. New York: Harcourt, 1991.
- Machiavelli, Niccolo. *The Discourses*. New York: Penguin, 1984.
- Michelman, Frank. “Law's Republic.” *Yale Law Journal* 97 (1988): 1493–1538.
- Pettit, Philip. *Republicanism: A Theory of Freedom and Government*. Cambridge, U.K.: Cambridge University Press, 1997.

- Pocock, J. G. A. *The Machiavellian Moment: Florentine Political Thought and the Atlantic Republican Tradition*. Princeton, NJ: Princeton University Press, 1975.
- Rawls, John. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1999.
- Rogers, Daniel T. “Republicanism: The Career of a Concept.” *The Journal of American History* 11 (1992): 11–38.
- Sandel, Michael J. *Democracy's Discontent: America's Search for a Public Philosophy*. Cambridge, MA: Harvard University Press, 1996.
- Skinner, Quentin. “The Idea of Negative Liberty: Philosophical and Historical Perspectives.” In *Philosophy in History: Essays on the Historiography of Philosophy*, edited by Richard Rorty, J. B. Schneewind, and Quentin Skinner, 193–221. Cambridge, U.K.: Cambridge University Press, 1984.
- Skinner, Quentin. *Liberty Before Liberalism*. Cambridge, U.K.: Cambridge University Press, 1998.
- Skinner, Quentin, ed. *Machiavelli and Republicanism*. Cambridge, U.K.: Cambridge University Press, 1990.
- Sunstein, Cass. “Beyond the Republican Revival.” *Yale Law Journal* 97 (1988): 1539–1590.
- Wood, Gordon. *The Creation of the American Republic*. Chapel Hill: University of North Carolina Press, 1998.

*Paul Weithman (2005)*

## RESCHER, NICHOLAS

(1928–)

Born in Hagen, Germany, where his father had established a law practice after serving as a German army officer in the First World War (1939–1945), Rescher's family emigrated to the United States in 1938, and he was educated there, receiving his PhD from Princeton University in 1951 at the age of 22. Since 1961 he has taught at the University of Pittsburgh, where he serves as University Professor of Philosophy and also as vice chairman of the Center for the Philosophy of Science. He has published more than 300 articles in scholarly journals, has contributed to many encyclopedias and reference works, and has written more than 100 books in various areas of philosophy, including epistemology, metaphysics, value theory and social philosophy, logic, the philosophy of science, and the history of logic.

In various publications Rescher has developed a detailed and systematic theory on the nature and limits of human knowledge along with its central implications for metaphysics and for the theory of values and ethics broadly conceived. Best viewed as an analytic pragmatist, Rescher has sought primarily in many books and essays to revive and refurbish the idealistic tradition in epistemology and metaphysics. Although he has written extensively on metaethics and issues of value and justice, his primary

efforts in epistemology and metaphysics constitute the central focus, and his approach to philosophy is comprehensively expressed in a trilogy titled *A System of Pragmatic Idealism*.

Generally, Rescher affirms the centrality of the natural sciences as the privileged source of understanding the nature of the empirical world and as directing our actions within it. He emphasizes, however, that the presuppositions of the natural sciences cannot be directly defended in natural science without circularity; such presuppositions, therefore, fall into the realm of metaphysics and are to be defended philosophically, philosophy being distinct from, but inextricably dependent on the deliverances of, natural science.

Rescher also sees the scientific method(s) as the product of an evolutionary process of rational selection, which leaves us with only those methods that have been proven to work by way of providing reasonably precise predictions of our sensory experience. In short, the methods of natural science, as well as their presuppositions, find their justification ultimately in the fact that we have a deep need for the products of natural science and epistemology, thereby underscoring the deeply practical or pragmatic nature of the whole of the cognitive enterprise, and whatever theoretical conclusions we reach therein (Rescher 1992–1994, 2001).

Regarding foundational beliefs or basic knowledge, Rescher affirms that basic beliefs, like all factual beliefs, are fallible and hence subject to revision in the light of ongoing evidence. Such beliefs begin as working presumptions about how things generally are, and are accepted as true until experience requires their rejection, but until experience forces such rejection they qualify for acceptance as items of human knowledge and serve as evidence for other beliefs, nonbasic beliefs. On the question of nonbasic knowledge, or scientific knowledge, he has consistently argued in *Methodological Pragmatism* and elsewhere that while particular scientific theses established by the inductive methods of science may be false (although we must presume them to be true when strongly confirmed), rationality requires us to use such a method because they generally tend to produce more effectively supplementable beliefs about the physical world than any other methods available.

Rescher construes truth in terms of any classical formulation of the correspondence theory of truth satisfying Alfred Tarski (1902–1983) biconditionals, and he argues that the criterion for it is fully warranted, assertible belief. The satisfaction of this criterion in any given case, however, does not entail logically that the proposition is true

rather than our best estimate of, or approximation to, truth; and it would be irrational to ask (as skeptics do) for anything more in the pursuit of truth, for nothing more can be had (Rescher 2003).

At no time, then, can we be sure of having the truth, rather than a reliable, but fallible, estimate of how things are, and it is this essentially fallibilistic conclusion that leads to Rescher's antirealistic view that we cannot be sure at any given time whether science actually succeeds in correctly describing an external world, although indeed we have good reason in this fallibilism to suppose that there is an external world. And this same fallibilism leads, with the support of various arguments, to the essential incompleteness of our knowledge of the world (Rescher 1978, 1999, 2000a, 2001, 2003). His idealism is consciously not an idealism affirming that all properties are linguistic in nature, but it does emphasize the fact that all systems of knowledge are the products of pervasive and profound human cognitive construction (Rescher 2001).

By way of philosophical methodology Rescher adopts a view he calls philosophical standardism. He thinks, for example, that human knowledge is fundamentally and standardly a matter of justified true belief. Prevalent counterexamples to the classical definition of knowledge as justified true belief are maximally distortive of the fact that philosophical explanations are based on limited generalizations that are subject to revision and we seek what is normally and typically the case rather than what is unexceptionally and necessarily the case (Rescher 1994, 2003). For Rescher, then, traditional philosophy is too much given over to abstract necessities of general principle that do not capture our understanding of the world as it is actually experienced, and the price we pay for his more modest construal of philosophical generalizations is to acknowledge the essential open-endedness of our philosophically relevant concepts.

By way of compensation for this less demanding view of philosophical methodology, Rescher urges that we can resolve a host of philosophical problems, such as the Gettier problem that have lingered too long because of the mistaken and pervasive belief that philosophical generalizations will be adequate only if they do not admit of exception in any context (Rescher 1994, 2003). Rescher argues, then, for a different view of conceptual analysis, a view allowing us to resolve the Gettier problem as well as the problem about the concept of meaningfulness in empiricism. For example, he argues that the classical definition of knowledge in terms of justified true belief has led to a hopeless set of counterexamples and counterdefinitions simply because people think mistakenly that

counterexamples refute a concept defined, and this mistaken belief roots in a faulty concept of analysis that is traditional and aprioristic.

Rescher's proposal is that we construe the relationship between knowledge and justified true belief not as a definition, but as a merely standardistic or generalized linkage under which "Standardly, knowledge is justified true belief" is a perfectly acceptable generalization, not only plausible but largely unproblematic. For Rescher, in the context of an epistemological standardism, interpreting such generalizations in a standardistic way does not allow the definition to be annihilated by counterexample. After all, as he says, knowledge is pretty standardly justified true belief. This same approach he applies to the empiricist criterion of meaning.

Although Rescher ascribes a certain primacy to induction and the methods of natural science because they are the products of the evolutionary process, he has not argued that the only legitimately answerable questions are those that admit of answer under the methods of science. He in fact has argued against that view when, among other arguments, he defends metaphysics as that philosophical venture seeking to examine and criticize the presuppositions of natural science, which natural science cannot do without viciously circular reasoning. He also has claimed that such presuppositions find their ultimate justification in the ultimate consequences, formal and material, of accepting them and their capacity to satisfy human needs for practical adaption.

On the question of scientific progress Rescher has aggressively argued in various places that unto eternity science is progressive and revolutionary, meaning thereby that there will never be a time when we would be justified in believing that we had answered all answerable questions about the world. Owing to an inevitable exponential decay in our economic capacity to fund scientific technology, scientific progress will accordingly slow, without stopping, to increasingly infrequent theoretical and factual advances. But it will always be an open-ended and unfinishable affair (Rescher 1978, 2000a, 2000b, 2003).

With regard to scientific realism, Rescher advances a cautious form of scientific instrumentalism without endorsing instrumentalism as a whole on the issue of factual knowledge. For Rescher, commonsense beliefs (those beliefs so obviously true that we cannot even imagine factual conditions under which they would be false) do succeed in standardly describing the physical world because such beliefs are not in any way likely to suffer truth value revision (Rescher 2003). Scientific beliefs, however, have no such property and must, for that reason, be regarded

as instrumentally reliable beliefs that we can plausibly presume true when strongly confirmed.

Otherwise, Rescher's fundamental metaphysical view on the question of reality originates in what he calls his *pragmatic idealism*, which he also sees as an antirealist implication of fallibilism and is an idealism only to the extent that it emphasizes the constructive role of cognitive processes in structuring our beliefs about an external and independent world about which we have knowledge in terms of our capacity to estimate the truth in the light of available evidence and in terms of what we can reasonably ascertain as typically and generally the case. But, it is not an idealism denying the existence of an external world. That such a world exists fundamentally roots in the essentially fallibilistic limitations and incompleteness of our knowledge of the world, as has been demonstrated time and again in the history of science and elsewhere (Rescher 1992–1994, 2000a, 2001, 2003).

In several cases philosophers have associated Rescher's name with a particular concept or principle broadly discussed, most notably in *Rescher's Law of Logarithmic Returns*, *The Rescher Quantifier*, *Rescher's Effective Average Standard* in the theory of distributive justice, *The Dienes-Rescher Inference Engine* in nonstandard logic, and *The Rescher–Manor Mechanism* in non-monotonic reasoning theory.

**See also** Epistemology; Metaphysics; Social Epistemology.

## Bibliography

### PRIMARY WORKS

- Methodological Pragmatism: A Systems-Theoretic Approach to the Theory of Knowledge*. Oxford: Basil Blackwell, 1977.
- Scientific Progress: A Philosophical Essay on the Economics of Research in Natural Science*. Pittsburgh, PA: University of Pittsburgh Press, 1978.
- Scientific Realism: A Critical Appraisal*. Boston; Dordrecht, Netherlands: D. Reidel; Norwell, MA: Klumer Academic Publishers, 1987.
- A System of Pragmatic Idealism*. Princeton, NJ: Princeton University Press, 1992–1994.
- Philosophical Standardism: An Empiricist Approach to Philosophical Methodology*. Pittsburgh, PA: University of Pittsburgh Press, 1994.
- The Limits of Science*. 2nd ed. Pittsburgh, PA: University of Pittsburgh Press, 1999.
- Nature and Understanding: The Metaphysics and Method of Science*. Oxford: Clarendon Press, 2000a.
- Realistic Pragmatism: An Introduction to Pragmatic Philosophy*. Albany: State University of New York Press, 2000b.
- Cognitive Pragmatism: the Theory of Knowledge in Pragmatic Perspective*. Pittsburgh, PA: University of Pittsburgh Press, 2001.

*Epistemology: An Introduction to the Theory of Knowledge.*  
Albany: State University of New York Press, 2003.

**Robert Almeder (2005)**

## RESPECT

The ideas that people should be treated with respect and that individuals should respect themselves are important elements of everyday morality and moral philosophy. Some theories treat respect for persons as the basis of morality or the hallmark of a just society, while self-respect is often viewed as a core moral duty or something that social institutions must support. There is disagreement, however, about whether things other than persons, such as animals or the environment, are appropriate objects of respect.

Most generally, respect is acknowledgement of an object as having importance, worth, authority, status, or power. As its Latin root *respicere* (to look back) indicates, to respect something is to pay attention or give consideration to it. As the etymology also suggests, respect is responsive: the object is regarded as due, deserving, or rightly claiming acknowledgement. Respect can be an unmediated emotional response, but it typically involves a conception of certain forms of acknowledgement as appropriate in virtue of some feature of or fact about the object, which is the basis of respect. Respect thus differs from attitudes such as liking, which are based in the agent's interests. Respect also typically involves behaving in ways that show regard for the object or refraining from certain conduct out of respect for it. We can respect rules by obeying them, dangerous things by taking precautions, and authorities by deferring to them; but respect is commonly thought to involve appreciating the value of the object. Valuing respect can be akin to admiration, awe, or honor, but contrasts with valuing modes such as maximizing and using. We can respect things we do not approve of, but regarding something as worthless or irrelevant is incompatible with respecting it.

There are many types of respect. Consider the well-mannered respect children should show parents and teachers, the great respect one might have for accomplished or morally exemplary individuals, the just respect people demand for their rights, the wary respect a prudent hiker has for rugged backcountry, the pro forma respect of standing for the judge entering a courtroom, and the basic respect many believe we owe people simply as people. These can be understood in terms of Stephen Darwall's (1977) now-standard distinction between two

fundamentally different kinds of respect: recognition respect and appraisal respect.

Recognition respect is a disposition to take something appropriately into account in deliberations about action. A diversity of things, including laws, rights, hazards, opinions, social institutions and positions, nature, and people can be objects of different forms of recognition respect. What recognition respect involves in various cases depends on the reasons why objects of that sort should be taken into account. Recognition respect is a moral attitude if the object is regarded from a moral point of view, for example, as having moral worth or as morally constraining actions. By contrast, we have appraisal respect (which some call evaluative respect) only for people, either as persons or in some role or activity, or for their qualities or achievements. Like esteem, it is based on a positive assessment of an individual's merits and admits of degree; but whereas any valued feature can be a basis of esteem, appraisal/evaluative respect concerns the moral quality of an individual's character. In addition, some philosophers regard the feeling of reverential respect as a distinct third kind of respect.

Whereas everyday discourse tends to use "respect" in the evaluative sense, as thinking highly of someone, philosophical attention focuses chiefly on moral recognition respect for persons. Individuals can be owed recognition respect in virtue of their social position (for example, as an elder or judge); such respect involves conforming to conventions for appropriate behavior. However, respect for persons commonly means recognition respect that all persons are morally owed solely because they are persons, regardless of social positions or individual qualities.

The moral philosophy of Immanuel Kant contains the most influential discussions of respect for persons. Kant holds that all and only persons, by virtue of their rational autonomy, are "ends in themselves" and have a special, unconditional worth called "dignity." Respect is the only fitting response to dignity; consequently, we have a fundamental and absolutely binding moral obligation to respect persons as ends in themselves. Moreover, all persons are equal in dignity and moral status with other persons, so each has a right to respect from all as well as a duty to respect themselves. Kant expresses this idea in *Groundwork of the Metaphysics of Morals* (1785) in one version of the categorical imperative, which is the supreme principle of morality: "Act so that you treat humanity, whether in your own person or the person of any other, never simply as a means but always at the same time as an end." In *The Metaphysics of Morals* (1797) Kant

explicates specific ethical duties of respect for others and self-respect.

Kant's account prompts numerous questions. Is rational autonomy indeed what gives persons the unconditional claim to moral recognition respect? Some thinkers argue that this is too thin a view of what matters morally about persons. Are all humans owed respect? What about those who lack rationality, such as profoundly mentally disabled individuals or human fetuses and embryos? Must persons always be respected regardless of moral merit, or can recognition respect be forfeited, for example, by evildoing? Some contend that remorseless evildoers warrant no respect; others hold that while they deserve punishment, they must still be respected as persons. What attitudes and conduct express respect or disrespect for persons? Humiliation, coercion, and enslavement are quintessential forms of disrespect; what positive measures (e.g., helping others pursue their ends, listening to their points of view) does respect require? What does respect imply for issues such as assisted suicide, pornography, poverty, and political rights for cultural minorities? Theorists also ask whether respect for persons is the foundation of all other moral duties and rights or simply one important moral consideration among others, and whether non-Kantian ethical approaches such as utilitarianism can accommodate the idea that persons are unconditionally owed respect.

A rich debate concerns whether things other than persons, such as other living things or the natural environment, which are often valued merely as means serving human interests, have a moral status that demands respect. Some thinkers argue that the basis of morally required respect is wider than rationality and can be possessed by nonpersons. Others hold that there are levels of respect such that while persons are owed maximal respect, other things may be due a lower level of respect that nevertheless rules out certain treatment, such as destroying them for trivial reasons. Widespread acknowledgment of duties of respect to nonpersons could entail significant changes in many human activities, such as eating, land and energy use, and biomedical research.

Self-respect, important in its own right, involves due appreciation of one's morally significant worth: worth one has either as a person or in some position or activity (recognition self-respect), or worth earned through the quality of one's character and conduct (evaluative self-respect). Both kinds of self-respect include an engaged understanding of the implications of having worth for directing one's life and interacting with others. Respecting oneself contrasts with, among other things, servility,

acquiescence to disrespect, shamelessness, chronic irresponsibility, self-destruction, and self-contempt. Evaluative self-respect is distinguishable from self-esteem. The former involves regarding one's character and conduct as coming up to scratch; it is lost if one comes to regard oneself as morally intolerable. The latter is enhanced or diminished through believing that one has or lacks any highly prized quality.

Self-respect is regarded both as morally required and as essential to the individual's well-being. It is thus strong criticism to say that a person does what no self-respecting person would do or that a social institution undermines people's self-respect. For Kant, individuals have a moral duty to respect their equal dignity as persons and to do nothing that would degrade or disavow it. In *A Theory of Justice* (1971) John Rawls maintains that because the ability of individuals to respect themselves is significantly affected by their social and political circumstances and because self-respect is vital to individual well-being, justice requires that sociopolitical institutions support self-respect. Connections between self-respect and, for example, responsibility, self-identity, forgiveness, prostitution, oppression, and education are also of philosophical interest.

*See also* Kantian Ethics; Moral Sentiments; Rights.

### **Bibliography**

- Buss, Sarah. "Respect for Persons." *Canadian Journal of Philosophy* 29 (4) (1999): 517–550.
- Darwall, Stephen L. "Two Kinds of Respect." *Ethics* 88 (1) (1977): 36–49.
- Dillon, Robin S., ed. *Dignity, Character, and Self-Respect*. New York: Routledge, 1995.
- Hill, Thomas E., Jr. *Respect, Pluralism, and Justice: Kantian Perspectives*. New York: Oxford University Press, 2000.
- Hudson, Stephen D. "The Nature of Respect." *Social Theory and Practice* 6 (1) (1980): 69–90.
- Kant, Immanuel. *Groundwork of the Metaphysics of Morals* (1785). In *Immanuel Kant: Practical Philosophy*. Translated and edited by Mary Gregor. New York: Cambridge University Press, 1996.
- Kant, Immanuel. *The Metaphysics of Morals* (1797). In *Immanuel Kant: Practical Philosophy*. Translated and edited by Mary Gregor. New York: Cambridge University Press, 1996.
- Rawls, John. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1971.
- Taylor, Paul W. *Respect for Nature: A Theory of Environmental Ethics*. Princeton, NJ: Princeton University Press, 1986.

**Robin S. Dillon (2005)**

## RESPONSE-DEPENDENCE THEORIES

The term *response-dependent* was introduced by Mark Johnston (1989) for concepts, such as *red*, that support an a priori biconditional on roughly the following lines: “X is red if, and only if, X is such that it would look red under normal conditions.” Any concept of the intended kind will apply to something just in case the object has a property in virtue of which it would elicit a relevant response, on a par with the appearance of redness, under relevantly favorable conditions; it will be akin to the concept of a secondary quality, traditionally conceived. The response to be elicited will involve a cognitive impression, so that the object looks, seems, or presents itself in a certain manner. And the conditions under which that response is guaranteed will have to be capable of independent specification; they cannot be defined just as whatever conditions will provide the guarantee.

Response-dependent concepts in this sense are meant to contrast with response-independent concepts whose application to an object depends solely on the nature of that thing in itself, not on the cognitive impression that the object makes on human beings. As Crispin Wright (1992) has emphasized in ongoing reflections around the theme, there must be a sense in which the object is of the conceptualized kind because it elicits that response, and not (or not just) the other way around; there must be a sense in which an object is red because it normally looks red.

The interest of the notion of response-dependence lies in the prospect of illuminating the character of a variety of concepts: for example, concepts of an evaluative, affective, or aesthetic kind; concepts associated with practices such as praise and blame or intervening causally in the world; concepts that are anthropocentric in any such manner; or perhaps all concepts that are mastered ostensively, without reliance on prior definition.

There are two very different theories of response-dependence in the literature. The biconditional associated with response-dependence, so all sides assume, does not hold because people’s relevant cognitive impressions never miss or misrepresent anything. So what makes certain concepts response-dependent, assuming that some concepts are indeed of this kind? What underpins the truth of the biconditional that governs them? The two theories diverge on that question.

One theory, explored by Johnston himself, would say that certain concepts are response-dependent because the

properties they designate are dispositions in things to evoke the relevant responses. Under this account we use a term like *red* to apply to those things that are such as to look red in suitable conditions; we think of the property of redness as the higher-order property of things that have a lower-order property, maybe this, maybe that, which makes them look red in suitable conditions. According to this theory, the concepts are response-dependent because the properties are defined by reference to responses; the a priori biconditionals hold, because they reflect the character of the properties conceptualized.

This theory has the disadvantage that, as Johnston (1993, 1998) himself has argued, few of our concepts are response-dispositional in this sense. With concepts such as *red*, we want to say that something looks red because it is red, where this is a causal explanation. But it is not clear that that claim remains available if redness is construed as a disposition; looking red will be a manifestation of the disposition, not a contingent effect. The issue has been a focus of controversy (Menzies and Petit 1993, Miller 2001).

The alternative theory of response-dependence would avoid this difficulty (Jackson and Petit 2002, Petit 2002). While allowing that there may be response-dispositional concepts, it says that other concepts may be governed by an a priori biconditional, too (or, being partly response-dependent, by at least an “only if” conditional). That will not be because they are paired with anthropocentric dispositions, but because the explanation of why they are paired with their particular, response-independent referents is that those properties have certain anthropocentric effects. On this account *red* may refer, not to the disposition to look red, but to a perfectly physical property, such as a certain profile of surface spectral reflectances. The reason why it will refer to that property is that it is the one that elicits the appearances on which speakers rely in learning to use the term. And so a connection will remain in place between the presence of the property and the looks-red response. Response-dependence will become salient, not at the level of semantics where we pair off terms with items in the world but, to invoke a distinction made by Robert Stalnaker (2004), at the level of meta-semantics where we try to explain the pairings that obtain between words and world.

Suppose that people generally rely on appearances in using the term *red*. Suppose that they intend to refer to a common, objective property in using the term; they are not content to go their idiosyncratic ways. And suppose

that because of that intention they seek to coordinate their usage, discounting some of the appearances of redness on which they divide. If this enterprise of coordination is to have objective significance, then there must be an objective reason for speakers to discount some appearances and not others; equivalently, there must be an objective reason to treat certain factors as perturbing or limiting influences on appearances. Why should speakers indict some influences as perturbing or limiting, then, but not others? According to this approach, it will be right to indict factors such that by privileging situations of usage where they are absent—by treating those conditions as favorable—speakers can optimally satisfy their joint intention to pick out the same objective property in things (Pettit 2002); the associated practice will do best in helping them to triangulate on a common, presumptively objective feature.

This approach will make it a priori, for any property such as redness that is available to be named in our language, that X counts as red if, and only if, it is such that it would look red under favorable conditions, with *favorable* defined by reference to the practice of discounting. The basis of the response-dependence will now lie, not in the nature of the property, but in the requirements that must be fulfilled for the property to deserve the name of *red*; that is why the biconditional is restricted to properties available to be named.

This theory of response-dependence allows us to say that while things may be conceptualized as red because of the associated appearances, still their redness is causally responsible for such appearances. It can mark out certain concepts as special, on a par with the concept of redness, particularly if the concepts are ineradicably response-dependent. And yet it can allow us at the same time to be realists about redness and similar properties, even holding that a term like *red* refers rigidly to an actual-world property (Haukioja 2001). If one wants to espouse response-dependence without too deep a revision of common sense, this is the way to go. If one aspires to be revisionary, the other theory of response-dependence is the better option.

*See also* Ethical Naturalism; Metaethics; Philosophy of Language; Primary and Secondary Qualities; Semantics.

### **Bibliography**

- Haukioja, J. "The Modal Status of Basic Equations." *Philosophical Studies* 104 (2001): 115–122.
- Jackson, F., and P. Pettit. "Response-Dependence without Tears." *Philosophical Issues* 12 (2002): 97–117.

- Johnston, M. "Are Manifest Qualities Response-Dependent?" *Monist* 81 (1998): 3–43.
- Johnston, M. "Dispositional Theories of Value." *Proceedings of the Aristotelian Society* 63 (1989): 139–174.
- Johnston, M. "Objectivity Refigured: Pragmatism with Verificationism." In *Reality, Representation, and Projection*, edited by J. Haldane and C. Wright. Oxford: Oxford University Press, 1993.
- Menzies, P., and P. Petit. "Found: The Missing Explanation." *Analysis* 53 (1993): 100–109.
- Miller, A. "The Missing Explanation Argument Revisited." *Analysis* 61 (2001): 76–86.
- Petit, P. *Rules, Reasons, and Norms: Selected Essays*. Oxford: Oxford University Press, 2002.
- Stalnaker, R. "Conceptual Truth and Metaphysical Necessity." *Philosophical Studies* 18 (2004).
- Wright, C. *Truth and Objectivity*. Cambridge, MA: Harvard University Press, 1992.

*Philip Pettit (2005)*

## RESPONSIBILITY, MORAL AND LEGAL

The term *responsibility* or one of its variants figures in moral discussion in many different ways. Philosophers have traditionally been especially interested in the concept of moral or personal responsibility. It is with the problems connected with this notion that the following discussion is primarily concerned.

### JUDGMENTS OF PERSONAL RESPONSIBILITY

F. H. Bradley once claimed that "for practical purposes we need make no distinction between responsibility and liability to punishment." Although it is true that discussions of responsibility have often turned quickly to discussions of blameworthiness and liability to punishment, there is little justification for Bradley's claim. For responsibility is equally relevant to many other forms of social treatment—among others, praise, reward (including special honors such as honorary degrees or titles), legal punishment, legal liability. And, of course, the topic is intimately related to the theological issue of salvation, the allocation of divine rewards and punishments.

Judgments of personal responsibility pertain to this range of practices in a very special way. Unless a person is judged personally responsible for some act or outcome, he would not normally be thought to deserve blame, praise, reward, punishment, and so on. Personal responsibility is generally regarded as a necessary condition of the justice of a person's receiving what he deserves. Yet



Bradley's error is repeated in many contemporary discussions of "freedom and responsibility" that start with some unilluminating remarks about "responsibility," then move swiftly to examination of blame or punishment. Discussion of responsibility is theoretically fundamental, not ancillary, to accounts of such practices.

### MEANING OF "MORAL RESPONSIBILITY"

Persons are normally judged morally responsible for their actions. But they may be judged responsible for almost anything—events, processes, their own psychological characteristics. Thus, a person may be judged morally responsible for his firm's loss of a contract, the Napoleonic wars, his bad temper, a technique for maintaining the fertility of land, or his friend's divorce. Under what conditions is a person responsible for one of his acts or for some other occurrence? If we can state the necessary and sufficient conditions for judgments of moral responsibility, we shall, in the process, be assigning a sense to the expression. "Moral responsibility," like so many other terms of moral discourse, is inevitably defined persuasively, for one is bound to be influenced in defining it by convictions about the requirements for deserved blame, praise, and so on. That is, one is bound to be influenced by convictions, explicit or implicit, about the requirements of justice in such matters.

Most persons, however, would accept the following form of definition, although those with different moral outlooks would complete it differently: A person is regarded as morally responsible for some act or occurrence *x* if and only if he is believed (1) to have done *x*, or to have brought *x* about; and (2) to have done it or brought it about freely. The completion of this formulation depends on what is meant by a human action; what would count as bringing some outcome about; and, above all, in what sense the terms *free*, *freely*, or *freedom* are employed. All these conceptions are problematic in ways that lead to very different theories of responsibility. Philosophers have too often supposed that the concept of "freedom" essential to moral responsibility can be fixed independently of what it is to be responsible, and that only after the meaning of *freedom* is specified can we determine whether, and under what conditions, a person is responsible. But in fact what a person means by *free*, *freely*, or *freedom* will reflect his moral convictions, and especially his views about justice, in the same way and for the same reasons that his conception of "moral responsibility" will reflect these views. As Harald Ofstad put it, "Ethical systems may determine the sense of 'freedom' we

select as relevant" (*Freedom of Decision*, p. 279). One need add only that they not only may, they do.

### FREEDOM AND MORAL RESPONSIBILITY

In his *Nicomachean Ethics*, Aristotle tried to analyze the concept of "voluntary action." Nowhere in his discussion did he clearly take account of the problems that arise if all our decisions and actions are determined by circumstances beyond our control. But he did claim that actions are compulsory "when the cause is in the external circumstances and the agent contributes nothing." It is difficult to say whether, in this and other passages, Aristotle intended to claim that the fact that the cause of action is external implies that the agent contributes nothing and is therefore not free in the sense relevant to responsibility. But from the beginning of the Christian era, the view that if decisions and actions are so determined, then persons are not free in the relevant sense, has been forcefully advanced and denied by countless numbers of theologians and philosophers.

The earliest form of the controversy arose in the context of Christian doctrine. In particular the fourth-century Christian theologian Pelagius argued that the doctrines of original sin and grace, and of divine omnipotence and foreknowledge, led to morally repugnant conclusions, primarily the conclusion that although a person's tendencies, decisions, and actions are in no way the fault of the agent, he is nevertheless morally culpable for his actions and, in consequence, justly suffers the torments of hell. If these doctrines are true, Pelagius argued, God is not just. But as God is certainly just, these doctrines must be false. Pelagius insisted that man is possessed of free will in that he has the power of "contrary choice." This power makes it possible for men to sin. In the fifth century St. Augustine countered Pelagius's attack on orthodox doctrine with the claim that though God knows and wills all, he grants to each person who has faith freedom of choice. Though God knows what a man will do, he wills only hypothetical claims, of the form "If this man sins, then he *shall* be punished." Divine decrees of this kind are consistent with freedom of the will. But what about the possession of faith—is this in a man's power? St. Augustine insisted that it was; for to have faith is to believe, and "belief is simply consenting to the truth of what is said, and consent is necessarily an act of will. It follows that faith must be in our power."

Although the terms are often different, the issues generated by this exchange persist. The doctrine of scientific determinism, and not the doctrine of divine

omnipotence, is today more commonly thought to pose the chief difficulties. Scientific determinists maintain that external conditions specified in scientific laws are sufficient to produce each human choice and action. But the nature of free choice, the nature of human power and ability, the relevance of necessity to freedom, the role of choice and deliberation, the very possibility of human choice, and many other issues suggested by or actually crystallized in the debate between Pelagius and Augustine are still vigorously debated.

### DILEMMA OF MORAL RESPONSIBILITY

Efforts to solve the problem of freedom of the will are conveniently considered against the background of the following dilemma.

If determinism is true, then all events, including any person's decisions and actions, are fully determined by circumstances that are ultimately beyond that person's control. If this is so, then that person could not have decided or acted differently. Hence the person was not free.

If determinism is false, then there are at least some events that are not fully determined by antecedent circumstances. To the extent that human decisions and actions are among those events which are not fully determined, those decisions and actions occur by pure chance. But what occurs by pure chance is not within a person's control. Therefore, to the extent that decision and action are not determined, the person is unfree.

But determinism is either true or false. Hence a person is never free with respect to decisions, actions, or the results of actions.

But, it is claimed, a person is morally responsible for an action or occurrence only if he is free in that respect.

Therefore, no one is ever morally responsible for any decision, action, or outcome.

**FREEDOM AS THE LACK OF CONSTRAINT.** Some philosophers have argued that determinism does not imply that a person's actions are beyond the person's control. They argue that there is a perfectly clear, ordinary sense to "being able" or "being free" to do something that is compatible with determinism. As Jonathan Edwards, the great American theologian, put it in *Freedom of the Will*, the most sustained, penetrating defense of this position: "Let the person come by his volition or choice how he will, yet, if he is able, and there is nothing in the way to

hinder his pursuing and executing his will, the man is fully and perfectly free, according to the primary and common notion of freedom" (Paul Ramsey, ed., 1957, p. 164). The central assumption of Edwards's argument is that the ordinary sense of statements like "Eisenhower could have ordered his troops to take Berlin before the Russians arrived" and "Kennedy was able to call off the invasion of Cuba, but he decided not to do so" is such that these statements are perfectly consistent with determinism. In David Hume's terms, there is an important distinction between an action being caused or determined by antecedent circumstances, and its being constrained or compelled or coerced by antecedent circumstances. Only when an action that is determined is also in some way constrained or compelled is the actor not morally responsible for that act.

Other philosophers have found this position unacceptable for a variety of reasons. Some have argued that the ordinary use of such expressions as "was free to," "could have," and "was able to" involves more than lack of constraint. They argue that careful analysis reveals that determinism is indeed inconsistent with statements of the form "X could have done such-and-such." Others have argued that freedom of the will depends upon freedom of decision, not freedom of action; and that if decisions are determined then it surely cannot be the case that one could have decided other than he did. Still others have claimed that there is no reason to accept the authority of common sense or ordinary language in these matters; that it is the philosopher's job to subject our common opinions to the test of careful, reasoned scrutiny, in the manner of Socrates.

**MORAL JUDGMENTS AND RESPONSIBILITY.** Among those who reject common sense as reflected in ordinary language as a basis for philosophical opinion are those who nevertheless endorse the distinction between constrained and unconstrained causally determined actions but defend it on explicitly moral grounds. Thus, certain philosophers have argued that the aim of holding someone morally responsible should be to influence future behavior in desirable ways—that, indeed, moral responsibility consists in the ability to be influenced by moral judgments. If a judgment of responsibility will not affect behavior in desirable ways, then there is no moral point in holding that person responsible. On this view, most customary excuses will still be acceptable. For it will not, in general, be possible to exert beneficial influence on a person if he did what he did either unintentionally or because no other course of action was possible.

One difficulty with this position is that we are, after all, concerned with persons other than the one whose responsibility is being judged. This concern can be accommodated by taking into account all of the consequences of a given judgment of moral responsibility, and determining whether the consequences are good, or best on the whole. But such a position seems to imply that a person believed to be innocent of an offense might be held morally responsible and be blamed or convicted on the general grounds that it would be socially beneficial to do so. And this seems to conflict with deeply held convictions about the requirements of justice in our commerce with other human beings. Considerations of this sort led Immanuel Kant to warn against the “serpent-windings” of utilitarianism. Utilitarianism seems to many to imply just such an unqualified appeal to social consequences.

Many thinkers feel that a related consideration has great importance in assigning moral responsibility. They have argued that the claims of justice are satisfied if we justify the rules according to which a person is judged to be morally responsible and blameworthy on the basis of the principle that social utility ought to be maximized, but then apply these rules to particular cases in a way that precludes any further appeal to this principle of utility. In this way, the claims of justice may be satisfied and the problem of freedom bypassed. This view, usually called “rule utilitarianism,” has been vigorously discussed by many contemporary moral philosophers. One criticism of it is that the restriction placed on the relevance of the principle of utility cannot itself be justified on utilitarian grounds, and that therefore the principles of justice cannot be explained or defended on a purely utilitarian basis.

**FREEDOM AS SELF-DETERMINATION.** Another gambit directed against the first argument of the dilemma rests on the distinction between self-determined action and action determined by circumstances external to the agent. Thus, Bradley argued that it is the self that may determine action and that, to the extent that this is so, the person is morally responsible for his actions. He argued that self-determinism does not imply that actions are predictable; actions are, in fact, not predictable, provided that the determining conditions are not entirely “materialistic” because they include “spiritual” or, perhaps, mental causes. The difference between the views of freedom as self-determination and freedom as absence of external constraint is that, although the latter allows that nonconstraining circumstances may be bodily causes external to the agent, the former view rules out this possibility. However, even if one could formulate a clear notion of the self that determines action, there seems to be no reason to

suppose that that self, or its determining characteristics, are themselves not determined by circumstances external to the agent. And if this is so, then the action would seem to be determined by circumstances that are ultimately beyond the person’s control. In reply to this objection it has been suggested that determinism does not imply that determinants occur before that which is determined—and that in the case of human decisions and action, the causal determinants occur simultaneously with the decision that in turn accounts for the action. Thus, the action is determined by a decision that is not itself the result of circumstances beyond the person’s control. For, as the determinants are concurrent conditions, in principle they can be affected by prior action. But it is not clear that this view rests on anything more than an ad hoc assumption needed to establish the possibility of self-determinism. There is, moreover, much psychological evidence for the view that if one’s decisions and actions are determined, then the determinants are circumstances temporally prior to them and external to the agent who decides and acts.

**INDETERMINISM.** Philosophers have been equally fertile in rebutting the second argument of the dilemma. Those who believe that only if determinism is false can a person be morally responsible, and thereby are impelled to attack this second argument, are usually called “libertarians” because they believe that the will itself is free in the sense of being undetermined. Libertarians claim that the fact that a decision or action is not fully determined by antecedent conditions does not imply that it occurred by “chance” or “accident” in a way that confers exemption from moral responsibility. But this argument does not refute the claim that an undetermined event is a matter of chance in a way that implies that it occurred by chance or by accident, in the sense of those terms that *is* relevant to moral responsibility. For example, the difference between knocking a flowerpot off a shelf as the result of the fully determined but accidental motion of someone’s arm or as a result of an undetermined motion of that arm seems irrelevant to a judgment of responsibility. The two events seem equally to void the responsibility of the agent. Both occurrences seem *accidental* in the relevant sense.

Others—J. D. Mabbott, for instance—claim that the first argument of the dilemma is sound, but it is inconceivable that moral responsibility is inapplicable to the human situation and, therefore, the second argument of the dilemma must be unsound. However, this is hardly an argument; it is rather a dogmatic affirmation of the point at issue. Still another argument is that human beings are so constituted that they necessarily hold others responsi-

ble for their actions and necessarily employ concepts in doing so that presuppose indeterminism. This conclusion would seem to rest on dubious psychological assumptions. In any event, if one could develop an account of moral responsibility that does not presuppose that determinism is false, which is morally defensible, and acceptance of which is psychologically possible, this view would be refuted.

**HARD DETERMINISM.** There have been other ingenious efforts to escape the toils of the dilemma. But it has also been argued that persons are indeed never morally responsible. According to this view, which has been called “hard determinism,” determinism is true and the first argument of the dilemma is sound. Hard determinists allow that blame and punishment may be useful, but they deny that they are ever morally deserved. Persons who blame or punish should do so only when engaged in moral education or social engineering; and blame and punishment have no special moral significance when they are justified as effective aids in these tasks. As the blame is not moral blame, there is no need to establish that it is deserved in virtue of the fact that the person is morally responsible. As John Hospers put it: “When we view other people’s frailties and shortcomings in the light of this perspective, we shall no longer say, ‘He deserves what he’s getting.’ Instead, we shall say, ‘There, but for the grace of God (and a favorable early environment) go I’” (*Human Conduct*, p. 521).

Hard determinists forget, however, that the claim that someone deserves what he is getting is not necessarily an expression of moral indignation. It may instead be an expression of the belief that all of the requirements of justice have been satisfied. If it is defensible to suppose that “freedom,” used in some sense consistent with determinism, is a requirement of justice, then hard determinism is unacceptable.

The general defect of the dilemma is that it presupposes that the relevant sense of “freedom” can be specified independently of a specific moral outlook, and particularly of a conception of justice. This defect reverses the proper order of moral reflection. The sense in which one can be said to have “acted freely,” and therefore to be morally or personally responsible and to deserve blame or praise or punishment or reward, should be specified in the light of one’s moral outlook—not independently of it.

## LEGAL RESPONSIBILITY AND PUNISHMENT

Many philosophers regard the legal context as paradigmatic for the discussion of moral responsibility. It seems clear that the unfortunate tendency to identify moral responsibility with blame and punishment derives partly from this fact. Nevertheless, the assessment of legal responsibility is so closely related to the assessment of moral responsibility, and legal experts have given such sustained and imaginative attention to the task of articulating criteria that are applicable to complex cases, that a careful study of the relevant aspects of the law will certainly assist the development of an adequate account of moral responsibility. Though problems pertaining to responsibility occur in all branches of the law, criminal law has received the most attention; the topics most frequently discussed in this connection are *mens rea* and criminal insanity.

**MENS REA.** The doctrine of *mens rea* requires a certain “mental element” to have been present when the offense was committed. This mental element is usually, but misleadingly, described as “guilty mind.” The characterization is misleading, first, because it is generally supposed that the offender need not be aware that he is committing an offense (“ignorance of the law is no excuse”); and, second, because many advocates of *mens rea* do not even require that the offender be morally culpable. On this second point there is, in fact, considerable disagreement. Some argue that unless an offender is morally blameworthy for his offense, he does not deserve to be convicted. Others insist on the distinction between moral responsibility and moral blameworthiness, arguing that a person may be morally responsible and may deserve to be convicted and punished for a crime even though his actions were not blameworthy. Broadly speaking, then, those who subscribe to the doctrine of *mens rea* believe at least that only persons who are morally responsible for their offense deserve conviction and punishment.

Discussions of *mens rea* usually take for granted the possibility of resolving the philosophical perplexities described above. Certain assumptions, generally unexamined, are made, and the work of articulating criteria appropriate to the criminal law goes forward. Those who accept the doctrine of *mens rea* in any of its forms believe that the requirement is satisfied if the offender has committed his offense intentionally. Some also claim that unintentional actions that are performed recklessly or negligently involve the necessary mental element. In general, the person who commits an offense is thought to

have satisfied the doctrine of mens rea if he knew what he was doing at the time or if he would have known what he was doing had he proceeded with reasonable care and deliberation. The extent to which an offender is able to or actually does exercise deliberate control over his actions and their results seems to be central to the way in which moral responsibility as a condition of deserved conviction and punishment is incorporated into the criminal law. This point is, however, more general than the doctrine of mens rea itself—it being possible for someone to have acted intentionally while, by reason of mental defect, not possessing deliberate control over his actions. Before going on to this point, two criticisms of the doctrine of mens rea should be considered.

**Objections to mens rea.** There are those who argue that, at least for certain criminal offenses, the requirement of mens rea ought to be abandoned and that strict liability ought to prevail. That is, for certain offenses it does not matter that the act was unintentional and it does not matter that reasonable care was taken. There are various arguments for strict liability, but, in general, the case for it is specific to the offense.

Though the agent's state of mind would seem not to enter into legal deliberations where strict liability prevails, this is not quite so. For example, it has been held that a bank director is strictly liable for borrowing money in excessive amounts from his own bank. In *State v. Lindberg*, 258 U.S. 250 (1922), the director pleaded that he had been assured that the money borrowed did not come from his own bank. Though the director did not borrow the money from his bank intentionally, the act of borrowing was itself intentional. A person cannot be said to have borrowed money that he accepted as a gift; his own intentions as well as the intentions of the donor are controlling. Though borrowing does, therefore, require a certain state of mind, the absence of the "mental element" involved in intentionally borrowing from one's own bank would be sufficient to discharge a person from moral responsibility. Insofar as the doctrine of mens rea is designed to satisfy the requirement that only a person who is morally responsible for some act or its result deserves to be held legally responsible and punished, strict liability conflicts with it.

**Criteria of mens rea.** The second criticism does not so much repudiate the requirement of mens rea in establishing responsibility as it criticizes the effort to develop criteria for mens rea. H. L. A. Hart argued that the practical meaning of mens rea is given in what is allowed as excuse or mitigation within the law. In order to determine whether mens rea is established, Hart argues, "it is neces-

sary to refer back to the various defenses; and then these general words (like 'mistake,' 'accident,' and so on) assume merely the status of convenient but sometimes misleading summaries expressing the absence of all the various conditions referring to the agents' knowledge or will which eliminate or reduce responsibility." In other words the general "rules" summarize accepted excuses, and there just are no general principles in terms of which we can account for the acceptance of specific excuses. Hart then generalizes his discussion of mens rea to pertain equally to the assessment of responsibility in nonlegal contexts.

This thesis encounters many difficulties. For one thing Hart neglects to distinguish adequately between exemption from responsibility and exemption from blame or legal responsibility. Thus, if a person defends himself against moral criticism of his having hit someone else by claiming that he was acting in self-defense, he is in effect accepting responsibility but rejecting blame on the grounds that he was justified in what he did. Second, if proposed as a purely descriptive thesis about our actual use of the language of "excuses," Hart's position begs the prescriptive claim that a general rationale of excuse and mitigation ought to be given—that otherwise the acceptance of a certain excuse is morally arbitrary. Those who defend mens rea try to meet this obligation by focusing on the element of awareness of what we are doing when we choose and act. Indeed, Hart becomes his own best critic when, in a later essay, he argues that the main rationale for excuse and mitigation within law is respect for "the claims of the individual as such, or at least as a *choosing being*."

**CRIMINAL INSANITY.** A person might intend to kill a particular person after careful deliberation, and do so; and this would be sufficient to satisfy mens rea. But if the offender suffered from extravagant delusions of having been persecuted by the person killed, he would normally be thought to be entitled to exemption from criminal liability on grounds of insanity.

The criterion of legal insanity generally adopted within Anglo American law is the M'Naghten Rule. This rule was formulated by the judges of England in 1843 in response to the public outcry that resulted when Daniel M'Naghten was acquitted, on grounds of criminal insanity, of murdering Sir Robert Peel's private secretary. M'Naghten had mistaken the secretary for Peel and had killed that unfortunate man while suffering from persecutory delusions about Peel's intentions toward him. The judges attempted to provide a morally sound, legally workable criterion for determining whether a person was

entitled to acquittal on grounds of criminal insanity. They affirmed that:

to establish a defence on the ground of insanity it must be clearly proved that, at the time of committing the act, the party accused was laboring under such a defect of *reason*, from disease of the mind, as not to know the *nature and quality of the act he was doing*, or, if he did know it, that he did not *know he was doing what was wrong*. ... The question has generally been, whether the accused at the time of doing the act *knew the difference between right and wrong*. [Italics added.]

The rule has been the object of vigorous attack and defense ever since its formulation. One type of criticism roughly follows the line of argument expressed by various parts of the dilemma formulated earlier. For example, Barbara Wootton, arguing from a determinist position, claimed that no acceptable criterion of criminal insanity can be formulated; that efforts to formulate an adequate criterion of mental defect, and, in the final analysis, of responsibility itself, shatter on the rock of the first argument of the dilemma. Consequently, all efforts to assess moral responsibility should be abandoned within the criminal law. The law should be concerned solely with treating the offender. It is clear that this "reform theory" approach to the criminal law would sweep away not only the insanity plea, but *mens rea* as well. Thomas Szasz, by contrast, argued that there is no such thing as a mental illness, that the insanity plea is never a valid excuse, and that, therefore, it ought to be abandoned. This argument leads to the same conclusion on policy as that reached by the reform theorists with respect to the insanity plea, but leaves *mens rea* intact. Szasz is not skeptical of moral responsibility as such. Others, like David Bazelon (in his Isaac Ray Award Lecture, "Equal Justice for the Unequal"), criticize the M'Naghten Rule as being too narrow—as not embracing all those defects of mind that entitle an offender to exemption on grounds of not having been morally responsible for his offense. It seems clear that many of the issues generated by this debate, as well as those that concern the doctrine of *mens rea*, await an adequate philosophical theory of moral responsibility.

## AN APPROACH TO A THEORY OF MORAL RESPONSIBILITY

An adequate theory of moral responsibility cannot identify moral responsibility with liability to blame or punishment. Moreover, any such theory must explicitly recognize what is, in any event, generally the case: that the

meaning assigned to the key concepts in the theory, particularly "freedom," reflects the moral outlook of its author.

The second point is of particular importance. Suppose one reflectively endorses a conception of justice according to which a person deserves blame or praise, reward or punishment, and so on, only if that person's decisions or actions are not determined. Then one should define "freedom" in such a way that "*P* decided (acted) freely" implies "*P*'s decision (action) was not determined." Correspondingly, suppose one endorses a conception of justice according to which a person deserves blame, and so on, only if his decisions or actions have some property that may or may not be causally determined by circumstances beyond his control. Then "freedom" ought to be defined in such a way that the meaning of "*P* decided (acted) freely" is consistent with determinism. It is our practical aims and interests that should govern the shape of our language, and not unreflected-upon linguistic habit that should govern the shape of our moral outlook.

Thus, a theory of justice is the essential foundation for a theory of moral responsibility. In this connection it should be remembered that just acts are not always right. (Would it be right to refrain from punishing an innocent person if the consequence was the destruction of human civilization?) Moreover, acts of blame, praise, reward, and punishment that are not just may sometimes be right. (One may be justified in blaming or praising an infant in order to influence his future behavior, but there would be no justice in it.)

**See also** Action; Aristotle; Augustine, St.; Bradley, Francis Herbert; Consequentialism; Determinism and Freedom; Edwards, Jonathan; Hart, Herbert Lionel Adolphus; Justice; Kant, Immanuel; Pelagius and Pelagianism; Philosophy of Law, History of; Punishment; Socrates; Utilitarianism.

## Bibliography

Some of the more interesting classical and contemporary discussions (in a vast literature) are cited below.

### GENERAL

For broad treatments of responsibility, see Harald Ofstad's *Freedom of Decision* (Oslo, 1960) and Austin Farrar's *The Freedom of the Will* (London: A. and C. Black, 1958). A number of important papers can be found in the anthologies *Free Will*, edited by Sidney Morgenbesser and James Walsh (Englewood Cliffs, NJ: Prentice-Hall, 1962); *Determinism and Freedom*, edited by Sidney Hook (New York: New York University Press, 1958); *Freedom and the Will*, edited by D. F. Pears (New York: St. Martin's Press,

1963); and *Freedom and Responsibility*, edited by Herbert Morris (Stanford, CA: Stanford University Press, 1961). The Morris collection contains many papers particularly relevant to legal responsibility and has an extensive bibliography. Historically, the most important work is Aristotle's *Nicomachean Ethics*.

#### DETERMINIST AND LIBERTARIAN DEFENSES

The best traditional defenses of the view that moral responsibility and determinism are compatible are Jonathan Edwards's *Freedom of the Will* (1754), edited by Paul Ramsey (New Haven, CT: Yale University Press, 1957), and David Hume's *An Enquiry concerning Human Understanding*, Ch. 8. Important contemporary statements of the same position have been made in C. L. Stevenson's *Ethics and Language* (New Haven, CT: Yale University Press, 1944), Ch. 14; Moritz Schlick's *Problems of Ethics* (Englewood Cliffs, NJ: Prentice-Hall, 1939), Ch. 7; and P. H. Nowell-Smith's *Ethics* (New York: Philosophical Library, 1957), Chs. 19–21.

The most influential recent defenses of the "libertarian" position are C. A. Campbell's "Is 'Free-Will' a Pseudo-Problem?," in *Mind* 60 (1951): 441–465, and Stuart Hampshire's *Thought and Action* (London: Chatto and Windus, 1959). See also J. D. Mabbott's "Free Will and Punishment," in *Contemporary British Philosophy*, edited by H. D. Lewis, 3rd series (London, 1956).

#### HARD DETERMINISM

For hard determinism, see Holbach's *Système de la nature* (London, 1770) and John Hospers, "Free Will and Psychoanalysis," in *A Modern Introduction to Philosophy*, edited by Paul Edwards and Arthur Pap, 2nd ed. (New York: Free Press, 1965), pp. 75–85. This volume contains an extensive bibliography of material bearing on the topic of the present entry on pp. 99–108.

#### ETHICAL THEORY

Some recent texts that place the problem of moral responsibility in the general context of an ethical theory are Richard Brandt, *Ethical Theory* (Englewood Cliffs, NJ: Prentice-Hall, 1959); A. C. Ewing, *Ethics* (London: English Universities Press, 1953); W. Frankena, *Ethics* (Englewood Cliffs, NJ: Prentice-Hall, 1963); and John Hospers, *Human Conduct* (New York: Harcourt Brace, 1961).

#### PROBLEMS IN THEORY-MAKING

Of the many technical articles dealing with various conceptual problems involved in the development of an adequate theory of moral responsibility, the following are of some interest.

On the difficulties of analyzing such terms as *can* and *ability*, see John Austin, "Ifs and Cans," in *Proceedings of the British Academy* 42 (1956): 109–132; Richard Taylor, "I Can," in *Philosophical Review* 59 (1960): 78–89, and A. S. Kaufman, "Ability," *Journal of Philosophy* 60 (September 12, 1963): 537–551.

On the difficulties of analyzing such terms as *intention* and *decision*, see Carl Ginet, "Can the Will Be Caused?," in *Philosophical Review* (1962): 49–52; Stuart Hampshire and H. L. A. Hart, "Decision, Intention, and Causality," in *Mind* 67 (1958): 1–12; G. E. M. Anscombe, *Intention* (Oxford: Blackwell, 1957), and A. S. Kaufman, "Practical Decision," in *Mind* 75 (1966): 25–44.

#### QUESTIONS IN CRIMINAL LAW

For works especially pertinent to the relevance of moral responsibility to the criminal law, see, in addition to the Morris collection (*Freedom and Responsibility*): J. D. J. Edwards, *Mens Rea in Statutory Offenses* (London, 1955); Joel Feinberg, "Problematic Responsibility in Law and Morals," *Philosophical Review* 71 (1962): 340–351; and G. L. Williams, *Criminal Law: The General Part* (London: Stevens, 1953), pp. 28–45, 77–81. See also Henry Hart's "The Aims of the Criminal Law," in *Law and Contemporary Problems* 23 (1958): 405–441, which presents a general defense of the claim that moral blameworthiness is an essential condition of criminal liability. For a contrary view, see H. L. A. Hart's "Legal Responsibility and Excuses," in Hook's collection, *Determinism and Freedom*; this article also in part constitutes an amendment to his own earlier essay, "The Ascription of Responsibility and Rights," *PAS* 59 (1949): 171–194, where he argues against the possibility of a general rationale of excuse and mitigation.

On the topic of strict liability, see R. A. Wasserstrom's "Strict Liability in the Criminal Law," *Stanford Law Review* 12 (1960): 730–745.

The literature dealing with criminal insanity is vast; see Barbara Wootton's *Social Science and Social Pathology* (London: Allen and Unwin, 1959), Ch. 8; Thomas Szasz's *Law, Liberty, and Psychiatry* (New York: Macmillan, 1963); David Bazelon's "Equal Justice for the Unequal" (The Isaac Ray Award Lecture). Unfortunately this last work is available only in mimeographed form.

Among scientific writings that have a special relevance are M. Gluckman, *The Judicial Process among the Barotse of Northern Rhodesia* (Manchester, U.K.: Manchester University Press, 1955), and B. F. Skinner, *Science and Human Behavior* (New York: Macmillan, 1953), Chs. 12 and 22.

*Arnold S. Kaufman (1967)*

## REVELATION

The notion of "revelation" is central to three of the major world religions: Judaism, Christianity, and Islam. Through Christianity in particular it has long been an important element in the religious thought of the West, and the present entry will treat it in this context, especially that of Christian theology.

During the twentieth century, but beginning in the nineteenth century, many—especially Protestant—theologians radically revised their conception of revelation. The view that was virtually axiomatic for all schools of thought in the mid-nineteenth century and that still remains the majority position (for it continues both in Roman Catholicism and in sections of conservative Protestantism) may be called the prepositional view of revelation.

## THE PROPOSITIONAL CONCEPT

In the propositional view, that which is revealed is a body of religious truths capable of being expressed in propositions. Because a knowledge of these truths is necessary for man's salvation, God has supernaturally made them known. Accordingly, in the words of the *Catholic Encyclopedia*, "Revelation may be defined as the communication of some truth by God to a rational creature through means which are beyond the ordinary course of nature" (Vol. XIII, p. 1).

The fuller significance of this propositional understanding of revelation appears when we view it in relation to three other basic theological categories with which it is closely connected. A particular conception of the nature of revelation involves a particular conception of the nature of faith, as man's response to revelation; of the Bible and its inspiration, as a medium of revelation; and of the character of theological thinking, as thought that proceeds on the basis of revelation.

When revelation is conceived as the divine disclosure of religious truths, faith is necessarily understood as the obedient believing of these truths. Thus faith was defined by the First Vatican Council (1870) as a supernatural virtue whereby "with the inspiration and help of God's grace, we believe that what he has revealed is true, not because its intrinsic truth is seen with the natural light of reason, but because of the authority of God who reveals it" (*Enchiridion Symbolorum*, edited by H. J. D. Denzinger, 29th ed., Freiburg, Germany, 1952, No. 1789).

The Bible finds its place in this system of thought as the book in which divinely imparted truths are written down and thereby made available to all humankind. Indeed, throughout considerable periods of Christian thought the Scriptures have been called the Word of God and have been virtually identified with revelation. The Bible is accordingly thought of as being ultimately of divine authorship; it has been written by human beings, but in the writing of it, their minds were directed by the Holy Spirit. Thus, the First Vatican Council said of the Scriptures that "because they were written as a result of the prompting of the Holy Spirit, they have God for their author" (*Deum habent auctorem*; Denzinger, *Enchiridion Symbolorum*, No. 1787); and in a similar vein, in the twentieth century, the Protestant evangelist Dr. Billy Graham said, "The Bible is a book written by God through thirty secretaries."

The propositional conception of revelation has also been integral to an understanding of the structure of theology that until recently has held unquestioned sway in

Christian thought since it was established by Thomas Aquinas in the thirteenth century. This hinges upon the distinction between natural and revealed theology. Natural theology comprises all those truths about God, and about the created universe in its relation to God, that can be arrived at by human reasoning without benefit of divine revelation. Accordingly, the core of natural theology consists in the traditional philosophical arguments for the existence of God. Revealed theology, on the other hand, comprises those truths about God, and about the created universe in its relation to God, that are not accessible to right reasoning as such and that can be known to men only because God has chosen to reveal them. (For example, it is held that while the existence of a supreme being is a tenet of natural theology, the further fact, stated in the Trinitarian dogma, that this being is "three Persons in one" belongs to revealed theology.) These various truths constitute the materials with which the theologian works, his primary task being to bring them together into a systematic body of doctrine.

These conceptions of faith, the Bible, and theology are linked together by the propositional character of revelation, with which they are all concerned. The revelation that is imparted by God, believed by men, published in the holy Scriptures, and systematized in the church's dogmas is a body of theological knowledge. This propositional conception of revelation began to form soon after the end of the New Testament period; reached its fullest development in medieval scholastic thought; was largely abandoned by the first Reformers in the sixteenth century, particularly Martin Luther, but became reestablished in the Protestant scholasticism of the seventeenth and eighteenth centuries; began to be questioned in the later nineteenth century; and was finally set aside by considerable sections of Protestant thought in the twentieth century.

## THE HEILSGESCHICHTLICH CONCEPTION

The fundamental premise of the propositional view has no place in the nonpropositional conception of revelation that was widely adopted by Christian theologians in the twentieth century. This view maintains that revelation consists not in the promulgation of divinely guaranteed truths but in the performance of self-revealing divine acts within human history. The locus of revelation is not propositions but events, and its content is not a body of truths about God but "the living God" revealing himself in his actions toward man. The nonpropositional view thus centers upon what has come in recent theology to be



known as *Heilsgeschichte* (salvation history) identified as the medium of revelation.

It is not supposed that God has marked his presence by performing a series of miracles, if “miracle” is taken to mean an event that compels a religious response by eluding all natural explanations. It is not characteristic of those theologians who think of revelation in nonpropositional terms to regard the biblical miracles as constituting theistic proofs. Rather, the *Heilsgeschichte* is the way in which a certain segment of human history—beginning with the origins of the national life of Israel and ending with the birth of the Christian community as a response to Jesus—was experienced by men of faith and became understood and remembered as the story of God’s gracious dealings with his people. What Christianity (and, confining itself to the Old Testament, Judaism) refer to as the story of salvation is a particular stream of history that was interpreted by prophets and apostles in the light of a profound and consistent ethical monotheism. They saw God at work around them in events that accordingly possessed revelatory significance. The *Heilsgeschichte* is thus a portion of history seen “from the inside” by the illumination of a particular religious faith. The publicly observable series of events forming its basis belongs to secular world history and is capable of a variety of political, economic, psychological, and other analyses besides that of theistic faith. As a central instance of this capacity of history to be construed both nonreligiously and religiously, Jesus of Nazareth, who has been seen by those outside the Christian community in various ways—for example, as rabbi, prophet, or political revolutionary—is seen by Christian faith as the divine Son incarnate in a human life, seeking to draw men into a new life in relation to God.

Revelation, understood in this way, presupposes faith as its correlate. That God is at work in a certain situation, which accordingly serves a revelatory purpose, is always a judgment of religious faith. The part played by faith is thus integral to the total event of revelation, if we use “revelation” to refer to the completed communication that occurs when God’s approach has met with a human response. In the words of William Temple, whose formulation of this conception of revelation has become classic, “there is event and appreciation; and in the coincidence of these the revelation consists” (*Nature, Man and God*, p. 314).

As in the case of its older rival, the fuller significance of what may be called the *heilsgeschichtlich* conception of revelation can best be indicated by sketching its implications for the understanding of faith, the Bible, and theo-

logical thinking. Clearly, in this view faith is not primarily the believing of revealed propositions, but is rather (in its cognitive aspect) a mode of discernment or interpretation in which men are convinced that they are conscious of God at work in and through certain events of both their personal experiences and world history.

The Bible is not a collection of divine oracles, but a record of the events through which God has revealed himself to a special group, a record that itself functions as a further medium of God’s self-revelation beyond that group. It has not been written at the dictation of the Holy Spirit, but has been composed by many different writers at different points within the period of the thousand years or so that it documents. It is distinguished from secular records of the same sequence of events by the fact that it is written throughout from the standpoint of faith. The Old Testament is dominated and unified by the God-centered interpretation of Hebrew history taught by the great prophets, in the light of which the story of the nation came to be understood and celebrated and its chronicles edited. The New Testament is dominated and unified by the witness of Jesus’ first disciples and of the Christian communities that grew up around them to the life, death, and resurrection of Jesus, whom they had received as the Christ. The faith by which alone the several writers could produce this particular literature constitutes the “inspiration” that has presided over its production.

Finally, there is no body of divinely authoritative theological propositions. Religious doctrines are not revealed, but represent human—and therefore fallible—attempts to understand the religious significance and implications of the revelatory events depicted in the Scriptures. Theologians who regard revelation in this manner have generally abandoned the traditional natural theology, with its theistic proofs, and base their doctrines instead upon faith as it responds to the scriptural records.

## SOME QUESTIONS

One of the questions that Christian theologians have repeatedly discussed is whether there is both general and special revelation. Are nature and history as a whole—including the whole religious history of humankind—revelatory of God, as well as the special occasions of the biblical *Heilsgeschichte*? Many theologians of all communions today hold that God is indeed universally active and that his activity always discloses something of his nature, even though his fullest personal self-revelation has occurred only in the person of Christ.

Another question that has at times been hotly disputed is whether there is an image of God (*imago dei*) in man that constitutes an innate capacity to respond to divine revelation (Emil Brunner) or whether, on the contrary, human nature is so totally corrupted by the Fall that in revealing himself to men God has to create in them a special capacity for response (Karl Barth).

The main philosophical question that arises concerns the criteria by which revelation claims may be judged. For proposition-centered religious thought the answer is provided by natural theology considered as a preamble to revelation. This establishes the existence of God and points, by means of miracles and fulfillments of ancient prophecy, to Christ and the Scriptures as the sources of revealed truth, supplemented in Roman Catholicism by the church as its divinely appointed guardian. For those theologies, on the other hand, that find God at work in historical events whose significance is discerned only by faith, there can be no proof of revelation. Such theologies arise within a community of faith (whether Jewish or Christian) that lives on the basis of what it believes to be an experience of divine revelation. It embodies in its life and literature the “memory” of momentous events in which God has opened a new and better life to humankind. The form of apologetic appropriate to this view is one that defends the right of the believer, as a rational being, given the distinctively religious experience out of which his faith has arisen, to trust that experience and to proceed to live upon the basis of it.

**See also** Barth, Karl; Brunner, Emil; Faith; Liberation Theology; Luther, Martin; Miracles; Thomas Aquinas, St.

### Bibliography

- The biblical basis of the idea of revelation is discussed in H. W. Robinson, *Inspiration and Revelation in the Old Testament* (Oxford: Clarendon Press, 1946; 4th ed., 1956), and E. F. Scott, *The New Testament Idea of Revelation* (London and New York: Scribners, 1935).
- R. Garrigou-Lagrange, *De Revelatione per Ecclesiam Catholicam Proposita*, 2 vols. (4th ed., Rome, 1945), is a classic Roman Catholic exposition. See also G. H. Joyce, “Revelation,” in *Catholic Encyclopedia*, edited by C. G. Herbermann and others, 15 vols. (New York: Appleton, 1907), Vol. XIII.
- John Baillie, *The Idea of Revelation in Recent Thought* (New York: Columbia University Press, 1956), traces the modern move in Protestantism to a nonpropositional conception of revelation. B. B. Warfield, *Revelation and Inspiration* (New York: Oxford University Press, 1927), provides a conservative, and C. H. Dodd’s *The Authority of the Bible* (London: Nisbet, 1928), a liberal, Protestant view of the place of the Bible in divine revelation. J. K. S. Reid’s *The*

*Authority of Scripture* (London: Methuen, 1957) represents a postliberal Protestant point of view.

The nonpropositional conception of revelation is expounded in William Temple, *Nature, Man and God* (London: Macmillan, 1935), Ch. 12; Emil Brunner, *Offenbarung und Vernunft* (Zürich, 1941), translated by Olive Wyon as *Revelation and Reason* (Philadelphia: Westminster Press, 1946); H. R. Niebuhr, *The Meaning of Revelation* (New York: Macmillan, 1941); Karl Barth, “The Christian Understanding of Revelation,” in *Against the Stream*, edited by R. G. Smith (New York: Philosophical Library, 1954); and H. Thielicke, *Offenbarung, Vernunft und Existenz* (Gütersloh, 1936). John Baillie and Hugh Martin, eds., *Revelation* (London: Faber and Faber, 1937), is a useful symposium.

For a contemporary Jewish discussion, see Abraham Heschel, *God in Search of Man* (New York: Meridian, 1956). For the Islamic conception of revelation, not touched upon in this entry, see I. Goldziher, *Die Richtungen der islamischen Koranauslegung* (Leiden: E.J. Brill, 1920), or A. J. Arberry, *Revelation and Reason in Islam* (New York: Macmillan, 1957); and for Hindu conceptions, K. S. Murty, *Revelation and Reason in Advaita Vedanta* (New York: Columbia University Press, 1959).

For contemporary skepticism concerning revelation, see Julian Huxley, *Religion without Revelation* (London: Watts, 1941; 2nd ed., London: Parrish, 1957), and Ronald Hepburn, *Christianity and Paradox* (London: Watts, 1958).

### OTHER RECOMMENDED TITLES

- Abraham, William. *Divine Revelation and the Limits of Historical Criticism*. Oxford: Oxford University Press, 1982.
- Avis, Paul, ed. *Divine Revelation*. Grand Rapids, MI: Eerdmans, 1997.
- Dulles, Avery. *Models of Revelation*. New York: Doubleday, 1983.
- Mavrodes, George. *Revelation in Religious Belief*. Philadelphia: Temple University Press, 1988.
- Swinburne, Richard. *Revelation: From Metaphor to Analogy*. Oxford: Clarendon Press, 1992.
- Trembath, Kern. *Divine Revelation: Our Moral Relation with God*. Oxford: Oxford University Press, 1991.
- Ward, Keith. *Religion and Revelation*. Oxford: Clarendon Press, 1994.
- Wolterstorff, Nicholas. *Divine Discourse: Philosophical Reflections on the Claim That God Speaks*. Cambridge, U.K.: Cambridge University Press, 1995.

**John Hick (1967)**

*Bibliography updated by Christian B. Miller (2005)*

## REVERSE MATHEMATICS

Reverse mathematics has its origins in Harvey Friedman’s 1974 address to the International Congress of Mathematicians. In it Friedman asked two fundamental questions: “What are the proper axioms to use in carrying out proofs of particular theorems, or bodies of theorems, in mathematics?” and “What are those formal systems which

isolate the essential properties needed to prove them?” Reverse mathematics was developed as an attempt to answer these questions, and since 1974 many logicians (especially Friedman and Stephen Simpson) have contributed to this project.

The goal in reverse mathematics is to find the minimal collection  $S$  of set theoretic axioms which suffices to prove a given theorem  $T$ . Because Zermelo-Frankel set theory is too powerful to provide this type of delicate analysis, second order arithmetic is used as the axiomatization of set theory. The formal language of second order arithmetic contains the symbols  $+$ ,  $\cdot$ ,  $<$ ,  $0$ ,  $1$ ,  $\in$ , and  $=$ , as well as two types of variables: number variables (denoted by lower case letters and intended to range over natural numbers) and set variables (denoted by upper case letters and intended to range over sets of natural numbers). In this formalization, sets of numbers are referred to using the set variables, but there are no variables that range over sets of sets of numbers. Thus, unlike Zermelo-Frankel set theory, second order arithmetic has to treat collections of sets as formal classes. One potential point of confusion concerning second order arithmetic is that despite its name, it is not a form of full second order logic. Second order arithmetic uses first order predicate logic, but allows two distinguished kinds of variables to separate its notation for numbers from its notation for sets. Therefore, the usual tools of first order logic such as compactness and the Lowenheim-Skolem theorems apply to second order arithmetic and its subsystems. In particular, there are countable models of second order arithmetic.

The axioms for second order arithmetic fall into three categories. First, there is a finite number of axioms stating the basic relationships between  $+$ ,  $\cdot$ ,  $<$ ,  $0$ , and  $1$  in the natural numbers. Two examples of these axioms are that for all  $m$ ,  $m + 0 = m$  and that for all  $m$  and  $n$ ,  $m \cdot (n + 1) = (m \cdot n) + m$ . Technically these axioms are exactly the noninduction axioms from Peano arithmetic. Second, there is an induction axiom for sets which says that from the assumptions that  $0$  is an element of  $X$  and that for all  $n$ , if  $n$  is an element of  $X$ , then  $n + 1$  is an element of  $X$ , we can conclude that every  $n$  is an element of  $X$ . This axiom captures the fundamental inductive nature of the natural numbers. Third, there is an infinite collection of axioms called the comprehension scheme. For each formula  $\varphi(x)$  in the language of second order arithmetic (allowing additional free variables as parameters), there is an axiom stating that there exists a set whose members are exactly the numbers  $n$  for which  $\varphi(n)$  holds. Because second order arithmetic does not allow the formation of sets of sets, these axioms do not give rise to a version of

Russell’s Paradox concerning the set of all sets which are not members of themselves.

The first step in analyzing a theorem  $T$  in reverse mathematics is to formalize the statement of  $T$  in second order arithmetic. This formalization can be done for most theorems in areas of mathematics that can be captured in some countable manner, such as classical geometry, number theory, real and complex analysis, countable algebra, and countable combinatorics. However, one of the limitations of using second order arithmetic as the underlying form of set theory is that it is not well suited to formalizing theorems from subjects such as general topology that depend heavily on the twentieth-century development of abstract set theory.

The second step is to find a subsystem  $S$  of second order arithmetic that is strong enough to prove  $T$ . This step often involves translating a classical proof of  $T$  into the formal system of second order arithmetic and letting  $S$  be the collection of axioms needed in the proof.

The third step is to show that  $T$  can prove each of the axioms in  $S$  over a suitably weak base theory (described below). This process of proving the axioms from the theorem is called a reversal, and it gives rise to the name reverse mathematics. If the system  $S$  contains axioms which are too powerful then the third step may be not possible. For example, a classical proof may use more axioms than are necessary. Therefore it is not uncommon to return to the second step and to try to find a different proof of  $T$  which uses weaker axioms. The third step is then repeated to see if these weaker axioms are provable from  $T$ . Once the third step is realized, the equivalence of the theorem  $T$  with the axioms  $S$  shows that  $S$  is a minimum collection of axioms which suffices to prove  $T$ .

Because the comprehension axioms (which state that for any formula  $\varphi(x)$ , there is a set whose elements are exactly the numbers  $n$  for which  $\varphi(n)$  holds) are the only ones which explicitly state the existence of sets, the subsystem  $S$  is often formed by restricting the types of formulas allowed in this scheme. For example, the base theory over which one typically proves the reversals is denoted RCA and is called Recursive Comprehension Axiom.

Roughly, RCA restricts the comprehension scheme to those formulas which define sets whose membership can be calculated by a finite algorithmic procedure.

Many of the theorems analyzed in reverse mathematics as of this writing fall into one of five categories. They are either provable in RCA or equivalent to one of four standard subsystems: WKL (which is RCA plus an

axiom stating that every infinite binary branching tree has an infinite path); ACA (which allows sets to be defined by formulas that do not contain quantifiers ranging over sets, but may contain quantifiers ranging over numbers); ATR (which allows sets to be defined by transfinite recursion); and  $\Pi_1^1\text{-CA}_0$ . (which allows sets to be defined by formulas containing arbitrarily many quantifiers ranging over numbers and at most one quantifier ranging over sets). A small number of examples do not fit neatly into these systems, but there is little evidence that this number will not grow as more theorems are analyzed.

One of the philosophical applications of reverse mathematics is that it provides a general framework for showing the necessity of impredicative methods in particular areas of mathematics. (Roughly, a set  $A$  is predicative if it can be defined as the set of all numbers satisfying a predicate for which the truth value does not depend on the existence of  $A$ .) On the one hand, if a theorem  $T$  is equivalent to a set of axioms that contains impredicative axioms (such as  $\Pi_1^1\text{-CA}_0$ .) then  $T$  cannot be established without the use of impredicative methods. On the other hand, if  $T$  can be proved inside a system which contains only predicative axioms (such as ACA), then this proof shows that  $T$  constitutes a piece of predicative mathematics. Some of the other subsystems have similar foundational connections. For example, because RCA restricts comprehension to formulas defining sets whose membership can be decided by a finite algorithmic procedure, proofs in RCA have a number of similarities to constructive proofs. However, this analogy is not perfect because most varieties of constructivism differ sharply with RCA over the treatment of induction and over the law of excluded middle.

A second philosophical application of reverse mathematics is to give a partial realization of Hilbert's program. In order to eliminate concerns over set theoretic paradoxes and to establish the consistency of infinitary methods in mathematics, Hilbert tried to find two formal systems to capture mathematical reasoning. The first system would be foundationally secure but would only capture finitary reasoning. The second system would be large enough to encompass all of mathematics including the general methods of infinitary reasoning. His goal was to use the first system to show the consistency of the second system, thus justifying the use of infinitary methods once and for all. Gödel's Second Incompleteness Theorem is widely viewed as showing that Hilbert's program cannot succeed as it was originally conceived. However, it is still possible that a reasonable fragment of mathematics could

be developed within a formal system that could be shown to be relatively consistent in a finitistic formal system. Hilbert did not provide a strict definition for his notion of finitary, but it has been argued that the system of primitive recursive arithmetic satisfies his intuitive concept. Furthermore, since primitive recursive arithmetic can prove the relative consistency of WKL, the (substantial amount of) mathematics that can be developed in WKL is finitely reducible in Hilbert's sense and provides a partial realization of Hilbert's program.

**See also** First-Order Logic; Geometry; Hilbert, David; Logic, History of: Modern Logic: From Frege to Gödel; Logic, History of: Modern Logic: Since Gödel: Friedman and Reverse Mathematics; Second-Order Logic; Set Theory.

### Bibliography

- Feferman, Solomon. "Systems of Predicative Analysis." *Journal of Symbolic Logic* 29 (1964): 1–30.
- Friedman, Harvey. "Some Systems of Second Order Arithmetic and Their Use." *Proceedings of the 1974 International Congress of Mathematicians. Volume 1*. Canadian Math Congress, 1975: 235–242.
- Simpson, Stephen G. "Partial Realizations of Hilbert's Program." *Journal of Symbolic Logic* 53 (1988): 349–363.
- Simpson, Stephen G. *Subsystems of Second Order Arithmetic*. Berlin: Springer, 1998.
- Tait, William W. "Finitism." *Journal of Philosophy* 78 (1981): 524–546.

*Peter Cholak (2005)*  
*Reed Solomon (2005)*

## RIBOT, THÉODULE ARMAND

(1839–1916)

Théodule Armand Ribot, the French psychologist, was a professor of psychology at the Sorbonne and from 1889 was the director of the psychological laboratory at the Collège de France. A philosophical disciple of Hippolyte Taine and Herbert Spencer (whose *Principles of Psychology* he translated), Ribot, with Taine, initiated the study in France of a positivistic and physiologically oriented psychology. His interest in philosophy was inseparable from his interest in concrete psychological problems and persisted throughout his life. He founded and edited the *Revue philosophique de la France et de l'étranger*, one of the first French philosophical journals. Ribot influenced not only French positivists and physiological psycholo-

gists but even some thinkers who, like Henri Bergson, rejected his epiphenomenalism.

Ribot's work falls into three main periods, but he remained loyal throughout his life to the program expounded in the introduction to his first book, *La psychologie anglaise contemporaine* (Paris, 1870). He insisted that psychology must be liberated from "the yoke of metaphysics" and stressed the need for an empirical, biological approach to psychology and the limitations of an exclusive reliance on introspection. However, although he insisted on excluding metaphysics from the empirical sciences, he did not dismiss it altogether. The works of Ribot's first period were mainly expository and historical. *La psychologie anglaise contemporaine* surveyed English associationist psychology from David Hartley to Samuel Bailey. In *La psychologie allemande contemporaine* (Paris, 1879) he introduced the work of Gustav Fechner, Wilhelm Wundt, Hermann Helmholtz, and others to the French public. *La philosophie de Schopenhauer* (Paris, 1874) foreshadowed Ribot's later emphasis on the affective and instinctive basis of personality.

Ribot's second period, characterized by an interest in psychopathology, produced three classic works: *Les maladies de la mémoire* (Paris, 1881), *Les maladies de la volonté* (Paris, 1883), and *Les maladies de la personnalité* (Paris, 1885). Despite a wealth of clinical, empirical material, the underlying motive of these works was philosophical—a positivistic distrust of such reified abstractions as "memory," "will," and "self." These abstractions had played a prominent role in French speculative psychology and in Victor Cousin's eclectic idealism. Ribot showed that the simplicity of such abstract words hides the complexity of the phenomenon named, a complexity revealed by the dissociation found in mental diseases. Ribot was among the first to study dissociations of personality, and his law of regression—that amnesia affects the most recent and least organized impressions and reactions first—was a lasting contribution to psychology.

In Ribot's third period, which began with his *La psychologie de l'attention* (Paris, 1888), his interest shifted to normal psychological phenomena, particularly to affective phenomena. The major work of this period, *La psychologie des sentiments* (Paris, 1896), reflects Ribot's biological approach and his epiphenomenalism. Physiological drives underlie our elementary feelings of pleasure and pain, and more complex and evolved stages of these drives underlie more complex emotions. Organic sensibility evolved prior to consciousness, and feelings prior to intellect. Ribot's last work, *La vie inconsciente et les mou-*

*vements* (Paris, 1914), interpreted various manifestations of subconscious activity in terms of motor activity.

**See also** Bergson, Henri; Fechner, Gustav Theodor; Hartley, David; Helmholtz, Hermann Ludwig von; Psychology; Spencer, Herbert; Taine, Hippolyte-Adolphe; Wundt, Wilhelm.

### **Bibliography**

- Dugas, Laurent. *La philosophie de Théodule Ribot*. Paris, 1924.  
 Faguet, Émile. *Propos littéraires*, fourth series. 1903.  
 Krauss, S. *Théodule Ribot's Psychologie*. Jena, 1905.  
 Lamarque, G. *Théodule Ribot: Choix des textes et étude de l'oeuvre*. Paris, 1913.  
 Taine, Hippolyte. *Derniers Essais de critique et d'histoire*, 6th ed. Paris: Hachette, 1923.

**Milič Čapek (1967)**

## RICHARD OF MEDIAVILLA

(d. c. 1300)

Richard of Mediavilla, or Richard of Middleton, *doctor solidus*, was a Franciscan philosopher, theologian, and canon lawyer. Although his date of birth and country of origin are unknown, scholars are generally agreed that he was either French or English. We are certain that in 1283 he was appointed as one of the judges of the works of Peter John Olivi, and we possess three of his sermons, preached in Paris in 1281 and 1283. He was a master of theology in Paris during 1284–1285. In 1288, Richard was one of the tutors of the exiled Prince Louis, son of King Charles II of Sicily and later bishop of Toulouse. Richard's last writings seem to date around 1295, when he completed his commentary on the fourth book of the *Sentences* of Peter Lombard. After 1295 we lose all trace of Richard of Mediavilla.

Richard was a scholar in the tradition of Bonaventure and John Peckham. He seems to have had a flair for clear and orderly presentation and to have enjoyed wide popularity among his Franciscan confreres. Like many of his fellow Franciscans, he regarded Bishop Tempier's condemnation of 219 propositions in 1277 as definitive. As a result, he set himself to defend, clarify, and organize a philosophy and theology that would vindicate and establish the doctrines contrary to the condemned propositions. He differs from most of his fellow Franciscans, however, in that he is more sympathetic to the Thomistic theory of knowledge.

Richard was one of the first Franciscans to reject the Augustinian theory of divine illumination. For Richard our ideas are solely the result of abstraction from sensible things, though as universals they are strictly intramental. In metaphysics he held that being is predicated analogically, not univocally, of God and creatures. Because every effect somehow bears the trademark of the first cause, God's existence can be proved from the world of nature. Richard found the so-called a priori argument of Anselm unacceptable; he adopted Henry of Ghent's position that essence and existence are only intentionally, not really, distinct. His doctrine of universal hylomorphism—that is, that all creatures are composed of matter and form—coincides with that of Bonaventure. Richard's theory of one substantial form's consisting of multiple grades constitutes the most complete and well-ordered doctrine of the plurality of forms in the Middle Ages. Richard argues to the soul's spirituality from the immateriality of universal concepts. The faculties of intellect and will are not accidents of the soul, nor do they add to its essence; they merely constitute a new relation between the essence of the soul and its acts and objects. Liberty is formally in the will. In common with his Franciscan confreres, Richard asserted that the will is a more noble faculty than the intellect.

Conservative by nature, Richard of Mediavilla was not one to shrink from speaking out. In one remarkable statement we catch a glimpse of his spirit in the search for truth and goodness: "We must start a good war. It is better to fight against falsehood and malice with a certain amount of discord, than, by dissimulating, to give way to malice and falsehood for the sake of harmony" (*Quodlibeta* III, 22).

**See also** Anselm, St.; Augustinianism; Bonaventure, St.; Henry of Ghent; Illumination; Medieval Philosophy; Olivi, Peter John; Peckham, John.

### **Bibliography**

#### WORKS BY RICHARD

- Commentarius in Sententias*, 4 vols. Brescia, 1591.  
*Questiones Disputatae*. Vaticani Latini No. 868. Folios 105–116.  
*Quodlibeta*. Brescia, 1591.

#### WORKS ON RICHARD

- Hocedez, Edgar. *Richard de Middleton. Sa vie, ses oeuvres, sa doctrine*. Louvain: "Spicilegium sacrum lovaniense" Bureaux, 1925.  
 Zavalloni, Roberto. *Richard de Mediavilla et la controverse sur la pluralité des formes*. Louvain, 1951. Has excellent bibliography.

*Girard J. Etzkorn (1967)*

## RICHARD OF MIDDLETON

See *Richard of Mediavilla*

## RICHARD OF SAINT VICTOR

See *Saint Victor, School of*

## RICKERT, HEINRICH

(1863–1936)

Heinrich Rickert, the German neo-Kantian philosopher, was born in Danzig and received his degree in 1888 from the University of Strasbourg. In 1891 he began lecturing at Freiburg, succeeding Alois Riehl as professor in 1894. In 1916 he went to Heidelberg as successor to Wilhelm Windelband.

Rickert belonged to the southwestern school of neo-Kantianism. His main efforts were devoted to a study of the logical and epistemological foundations of the natural sciences and to the historical disciplines in the hope of arriving at a "unity of reality and values." He departed from Wilhelm Dilthey in his criticism of Dilthey's subjective approach to the understanding of historical reality and in his attempt to find a set of more objective criteria; his departure from Windelband consisted in rejecting Windelband's separation of natural and historical disciplines and offering instead a theory that considered all reality to be historical.

### PHILOSOPHY AND NATURAL SCIENCE

In his early work, particularly in *Der Gegenstand der Erkenntnis* (Freiburg, 1892), Rickert raised the question of the relationship between philosophy and the natural sciences. He denied the universal validity of the method of the natural sciences and attempted to establish the primacy of practical reason as the foundation of his epistemology. He believed that only the Kantian critical method is adequate for explaining the epistemological presuppositions and limitations of the various sciences. While phenomenology may provide a method for describing the contents of consciousness, it fails to account for their intelligibility and relationship to objective reality. Hegelianism, on the other hand, in identifying the real with the rational, leaves out of account or distorts

the pluralistic character of reality. Only critical philosophy yields knowledge that is both universal and necessary; it alone can explain the pluralistic, dynamic, and yet rational character of society and history. In view of the lack of philosophical attention to the historical disciplines and because the then prominent philosophical problems of *Weltanschauung* seemed to hinge most directly on distinguishing scientific thinking from historical thinking, Rickert devoted himself thereafter primarily to the problem of historical conceptualization (*Begriffsbildung*).

### INDIVIDUALIZING AND GENERALIZING THOUGHT

On the basis of Windelband's distinction between nomothetic (universal) and ideographic (particular) judgments, Rickert developed his logic of the historical disciplines. At both the scientific and the prescientific stages of conceptualization, he claimed, there are two ways of grasping reality: individualizing and generalizing. Individualizing thought is proper to historical thinking. Instead of fabricating a copy of a historical phenomenon in its complex totality, it establishes the essential relationships that bind the phenomenon to its environment and traces the various stages of its development. Philosophy studies the concept of development, while the objects of historical study are unique developments. Generalizing thought, therefore, is proper to the natural sciences but is inapplicable to history. "Reality," Rickert claimed, "becomes nature if we consider it in regard to what is general; it becomes history if we consider it in regard to the particular or individual" (*Kulturwissenschaft und Naturwissenschaft*, 5th ed., p. 63).

Historical method for Rickert is highly selective, and in the selection of data, value judgments are operative from the very outset. This being the case, the determination of value criteria (*Wertbegriffe*) becomes the primary concern of historical understanding. Generalizing thought is logically free of values (*wertfrei*) because it constructs universally valid concepts. The particular objects to which they apply are interchangeable, and each object, abstracted from all its other relationships, functions only to illustrate the general law. Although in generalizing thought a selective process is at work to determine the common character of a group of particulars, it is the common character, expressed in a formula, that is essential. The aim of generalizing thought is precisely to free its objects from relations of value (*Wertverbindungen*).

### KULTURWISSENSCHAFT

Although history is a science of values, this does not mean that the historian may organize his inquiry arbitrarily; in that case history would be mere propaganda. In order for history to be objective, its values (state, law, art, religion) must be universal. The universality of historical values must be established epistemologically, and the relevance of the various social phenomena with respect to these values must be demonstrated empirically. Because history is written by, about, and for civilized men, social activity must be its subject matter. Since social activity can be grasped only by individualizing thought in terms of its significance for universal values, the historian's criterion must be culture, because social activity and value most nearly converge in culture. Culture is most directly concerned with the realization of universal values: "Culture is the common affair in the life of the nations; it is the possession with respect to the values of which the individuals sustain their significance in the recognition of all peoples, and the cultural values which adhere to this possession are therefore those which guide historical representation and conceptual formation in the selection of what is most essential" (*Die Grenzen der naturwissenschaftlichen Begriffsbildung*, 2nd ed., p. 509). Thus, believing that his method made of history a logically valid discipline that deals with objective reality, Rickert called the historical sciences *Kulturwissenschaft* (cultural science) in preference to Dilthey's term, *Geisteswissenschaft* (science of the mind or spirit).

### UNIVERSAL HISTORY

Far from being a contradiction, universal history is not only possible but is the logical outcome of the search for the value principles (*Wertprinzipien*) according to which the historical process as a whole may be viewed. "The system of values provides the possibility of systematization, and the relationship [of history] to the system of values permits of individualizing treatment" ("Geschichtsphilosophie," p. 400). But precisely because the evaluation of the whole of history is involved, the system of value principles must be purely formal. "We would need something timeless in order to extract an objective sense from the temporal course of history" (*ibid.*, p. 418). Like Immanuel Kant, Rickert proposed three stages in the development of civilization: dogmatism, skepticism, and criticism, the last of which was the achievement of German idealistic philosophy. While this periodization cannot be verified empirically, it is an example of the critical approach to the question of the unity of historical development. Although it is purely theoretical, it nonetheless

gives an axiological grounding to the results of empirical research. In the last analysis, the problem of universal history is to introduce a method whereby the real and the ideal may be theoretically synthesized.

## CRITICISM

The principal criticism brought against Rickert is that the introduction of a transcendental system of values is unhistorical and leads to the reification of existing values (*Wertabsolutierung*). In isolating universality by viewing it as a distinct realm of thought rather than as a function of all thought, Rickert actually confirmed the positivism and cultural relativism he had sought to overcome. In radically separating the universal from the particular, he was compelled to regard historical data as being identical with those of science, a series of discrete facts that differ only in the relationships in which they are observed. Nevertheless, the fruitfulness of Rickert's theory is borne out by his influence on such contemporaries as Ernst Troeltsch, Friedrich Meinecke, and Max Weber.

*See also* Dilthey, Wilhelm; Geisteswissenschaften; Hegelianism; Historicism; Meinecke, Friedrich; Neo-Kantianism; Phenomenology; Philosophy of History; Riehl, Alois; Troeltsch, Ernst; Weber, Max; Windelband, Wilhelm.

## Bibliography

### WORKS BY RICKERT

- Die Grenzen der naturwissenschaftlichen Begriffsbildung.* Tübingen: Mohr, 1896–1902; 2nd ed., Tübingen, 1913.  
*Kulturwissenschaft und Naturwissenschaft.* Tübingen, 1899; 5th ed., Tübingen, 1921.  
*Die Probleme der Geschichtsphilosophie.* Leipzig: Duncker and Humblot, 1905; 3rd ed., Heidelberg: C. Winter, 1924.  
 "Geschichtsphilosophie." In *Die Philosophie im Beginn des zwanzigsten Jahrhunderts, Festschrift für Kuno Fischer*, edited by Wilhelm Windelband. Heidelberg, 1907.  
*Kant als Philosoph der modernen Kultur.* Tübingen: Mohr, 1924.

### WORKS ON RICKERT

- Becher, E. *Geisteswissenschaften und Naturwissenschaften.* Munich: Duncker and Humblot, 1921.  
 Cohen, M. R. "The Insurgence against Reason." *Journal of Philosophy* 22 (1925): 120–123.  
 Collingwood, R. G., A. E. Taylor, and F. C. S. Schiller. "Are History and Science Different Kinds of Knowledge?" *Mind* 31 (1922): 426–466.  
 Faust, A. *Heinrich Rickert und seine Stellung innerhalb der deutschen Philosophie der Gegenwart.* Tübingen: Mohr, 1927.  
 Federici, F. *La filosofia dei valori di Heinrich Rickert.* Florence, 1933.

Gurvitch, Georges. "La théorie des valeurs de Heinrich Rickert." *Revue Philosophique* 124 (1937): 80–85.

Troeltsch, Ernst. "Über den Begriff einer historischen Dialektik. Windelband-Rickert und Hegel." *Historische Zeitschrift* 119 (1919): 373–426.

*Robert Anchor (1967)*

## RICOEUR, PAUL

(1913–2005)

Paul Ricoeur is widely regarded as among the most important French philosophers of the twentieth century. He had contributed to most of the major philosophical movements from the 1940s to the present, including existentialism, phenomenology, hermeneutics, structuralism, critical theory, narrative theory, philosophy of religion, ethical theory, political philosophy, and philosophy of law. Ricoeur was a prolific author of twenty-seven books and more than 500 articles as of 2004. His works tend to focus on theories of interpretation and the philosophy of human nature, examining the limits on our ability to understand the world and to know ourselves. If there is a guiding thread that runs through Ricoeur's career it would be an attempt to develop a philosophical anthropology of human capability, in particular our capacities to act, understand, communicate, and be responsible.

Born in 1913 in Valence, France, Ricoeur studied classics and philosophy at the University of Rennes and at the Sorbonne. After holding a number of teaching positions in provincial colleges, he was drafted into the French army in 1940. He was soon captured and spent the next five years in a German prison camp. While in prison, Ricoeur translated Edmund Husserl's book *Ideas* (1913) into French and coauthored a study on Karl Jaspers with fellow inmate Mikel Dufrenne. After he was freed in 1945, Ricoeur taught at the University of Strasbourg (1948–1956), the Sorbonne (1956–1966), and the University of Paris, Nanterre (1966–1987). In 1970 he succeeded Paul Tillich as the John Nuveen Professor of Philosophical Theology at the University of Chicago where he held a joint appointment at the School of Theology and Department of Philosophy until his retirement in 1992. Ricoeur continued to publish works on hermeneutics, moral-political philosophy, and theology until his death in May 2005.

## EXISTENTIAL-PHENOMENOLOGY

Ricoeur's early works were devoted to a phenomenological study of the human will. He sought to combine the



existentialist themes of Gabriel Marcel (incarnate existence) and Karl Jaspers (limit situations, such as birth, war, and death) with the methodological rigor of Husserlian phenomenology. The result is a proposed three-volume, systematic “philosophy of the will” that includes *Freedom and Nature: The Voluntary and the Involuntary* (1950), *Fallible Man* (1960), and *Symbolism of Evil* (1960). These works form the core of Ricoeur’s early philosophical anthropology. The third volume was to be on the “poetics of the will” but was never written.

In *Freedom and Nature*, Ricoeur employs the Husserlian method of eidetic analysis to the spheres of the will, affection, and volition. The goal is to describe the structures of voluntary action to uncover our fundamental possibilities of existence. Ricoeur retains from Husserl the central insight into the intentionality of consciousness and the methodological technique of bracketing, while recognizing that phenomenology must be supplemented with non-phenomenology given the limits placed on knowledge by the body. A phenomenology of action reveals the full extent to which consciousness is embodied and tied to involuntary functions, thus known both phenomenologically (as a subject of the will) and empirically (as an object for the will). The experience of our own bodies is never direct and unmediated; instead we interpret the involuntary aspects of our bodies as signs or symptoms for the will. These signs are read indirectly through one’s will as indications of the involuntary for the voluntary. We find consciousness in the body and the body in consciousness.

Ricoeur shows how the act of willing is both the realization of freedom and the reception of necessity. The act of willing has three moments, each inextricably related to the involuntary. The three parts are: 1) “I decide”; 2) “I move my body”; and 3) “I consent.” Each part has an object (or intentional correlate): A) the decision or project; B) the action or motion; C) the acquiescence or consent. Finally, each correlate is itself related to the different modes of the involuntary: a) motives, needs, values; b) skills, emotions, habits; c) character, unconscious, life. There is a fundamental reciprocity of voluntary decision, choice, and action with involuntary bodily functions, which act as a vehicle for the will. The involuntary necessity of the body both limits and enables human freedom. Yet the unity of mind and body, voluntary and involuntary, is never fully realized. Rather it is a regulative idea for understanding how humans are both free and constrained. Embodied freedom and unifying the will free from conflict is something we can only hope for but never completely realize. A poetics of the will in the proposed

third volume was to be Ricoeur’s attempt to show how imaginative and creative uses of language can suggest ways to reconcile the dualism of mind and body.

After having described the eidetic structures of the will as incarnate freedom, in the second volume of the philosophy of the will Ricoeur seeks to uncover the actual conditions of existence through an “empirics of the will.” One of Ricoeur’s aims is to overcome the tendency among existentialists to overvalue human transcendence and devalue human finitude. He believes that the mistake made by Jean-Paul Sartre and Martin Heidegger is to equate finitude (our inevitable and necessary limitations) with guilt (an undesirable experience of limitation). In *Fallible Man*, Ricoeur examines the conditions under which the will confronts its finitude and chooses evil and sin. The reason one chooses evil stems from the divided will. There is a disproportion between our finite limitations and infinite possibilities. This gap between our limited bios (our bodies, passions, and desires) and unlimited logos (our reason capable of grasping universals) renders us fragile and fallible. The fractured will, or fault in our existence (like a geological fault), opens the way for temptation, evil, and sin. We are not evil by nature but we have the capacity to be thanks to the disproportion in our will. Following Immanuel Kant, Ricoeur analyzes the existential significance of our fallibility in terms of imagination (the limits of knowledge), character (the limits of the will), and feeling (the limits of our emotion). The ineliminable conflict within human beings constitutes our capacity both for good and evil.

In the second volume of the empirics of the will, *The Symbolism of Evil*, Ricoeur continues to examine our capacity for evil by considering the various ways humans are already guilty, sinful, and fallen. He conducts a “phenomenology of confession” that describes the way we experience the transition from fallibility to fault (from our potential for evil to actually being evil). According to Ricoeur, confession arises from three sources: defilement, sin, and guilt. Defilement is interpreted as an objective state of impurity, sin as a social state, and guilt as a psychological interiorization of sin. Ricoeur then shows how this progressive “fallenness” is reproduced in four basic types of myth: myths of creation, myths of tragedy, myths of the fall, and myths of exile. Each type of myth is a symbolic expression of our experience of evil.

## HERMENEUTICS

Through his analysis of myths, Ricoeur began to shift away from phenomenology to hermeneutics as he became more interested in the symbolic systems that

relate us to the world and impose an indirect or interpretive approach to knowledge and self-understanding. Symbols are double-meaning expressions with an apparent, first-order, meaning and a hidden, second-order, meaning. Symbols must be interpreted rather than merely perceived in order to be understood. Ricoeur contends that if language is taken to be the medium for thought and experience, it is impossible to realize a pre-linguistic and presuppositionless realm of consciousness. As a result, we can never have the kind of unmediated knowledge that phenomenologists have traditionally hoped to attain. As such, the mediation of self-understanding by signs, symbols, and language requires an interpretive, hermeneutic philosophy. Ricoeur often speaks of the detour self-understanding must take through language. The idea of a detour as a hermeneutical technique for reading signs of experience through something else is one of Ricoeur's favorite metaphors that reappears throughout his career.

In *Freud and Philosophy* (1965), Ricoeur develops a hermeneutic philosophy by contrasting Husserlian phenomenology and Freudian psychoanalysis. According to Ricoeur, Sigmund Freud introduces a model for understanding the relationship between experience and desire, as well as a technique for uncovering the relation of a latent, unconscious meaning to a manifest, conscious meaning. The unconscious is an interplay of language and desire that reveals and conceals, thus shaping and distorting how we understand ourselves and others. Freud removes the illusion of a subject that ostensibly is immediately transparent to itself, thereby frustrating the aim of phenomenology to describe experience faithfully. For Ricoeur, Freud's contribution to hermeneutics is a theory of interpretation geared toward unmasking and decoding symbolic expressions. Dreams and symbols are models of the complexity of language in which meanings are both given and hidden. This symbolic language requires an interpretation in terms of rules and law-like regularities to understand how it mediates experience. Psychoanalysis, on this account, is a hermeneutic technique for interpreting the semantics of desire, that is, the interrelations among language, experience, bodily desires, and culture.

In *Freud and Philosophy*, Ricoeur contrasts two opposing kinds of hermeneutics: the hermeneutics of belief and the hermeneutics of suspicion. The hermeneutics of belief is geared toward recovering and recollecting lost or forgotten meanings. Understanding a religious symbol involves a hermeneutics of belief because to understand its full meaning one must already have the

prior belief that it is sacred. The hermeneutic situation is that we must believe in order to know, yet know in order to believe.

By contrast, the hermeneutics of suspicion is geared toward unmasking, demystifying, and removing the illusions of symbols, which not only reveal but conceal meaning. Ricoeur draws on the "masters of suspicion," Karl Marx, Friedrich Nietzsche and Freud, each of whom posit a false consciousness in place of an immediate, self-transparent consciousness, and deception or delusion in place of the experience of participation. The hermeneutics of suspicion decipher meanings hidden and distorted by literal and apparent meanings. Ricoeur argues that self-understanding involves a dialectic of belief and suspicion: We must have a clear understanding of our past that is shaped by a projection of what we hope we can become.

In *The Conflict of Interpretations* (1969), Ricoeur further develops a hermeneutic philosophy through his confrontation with structuralist semiotics. Like psychoanalysis, structuralist semiotics calls into question the primacy of consciousness as the privileged, self-evident home of meaning. Ricoeur retains the insight of structuralists, such as Ferdinand de Saussure and Claude Lévi-Strauss, that language has objective characteristics best understood as an empirical science and that meaning is a function of a different agency than consciousness. Yet Ricoeur maintains this aspect of language without rejecting the fundamental intentionality of consciousness and role of the individual as a bearer of meaning. Language has expressive meanings that must be understood from the perspective of the first person as well as objective meanings that must be understood from the perspective of the third person. Ricoeur tries to integrate a structuralist method of objective explanation into an interpretive theory for understanding spoken and written language.

As a result of the confrontation with structuralism, Ricoeur develops a theory of language as discourse. In *Interpretation Theory* (1971), discourse is defined as a dialectic of event and meaning, sense and reference. Discourse takes place as an event but has an ideal, repeatable meaning that allows what is said to be repeated, identified, and said differently. As an event, discourse is referential (about something), self-referential (said by someone), temporal (said at some moment), and communicative (said to someone).

As a meaning, discourse is both what the speaker means and what the utterance means. The dialectic of event and meaning, sense and reference constitutes writ-

ten discourse as well. But where spoken discourse is addressed to someone, in a particular dialogical situation, written discourse is addressed to an indefinite number of absent readers. The task of hermeneutics is to understand the matter of the text that is autonomous with respect to the intentions of the author, its original addressee, and the context in which it was written. The matter of the text discloses a proposed world (of real and imaginary references) one could possibly experience, inhabit, verify, criticize, and so on. To appropriate the meaning of a text we must first let go and relinquish the illusion that subjectivity alone confers meaning. Now the text, not the symbol, is Ricoeur's model for the linguistic mediation of experience. Self-interpretation is mediated by textual interpretation; conversely, textual interpretation results in self-interpretation.

The hermeneutics of texts also applies to actions. In his article *The Model of the Text as Meaningful Action* (1971), Ricoeur argues that actions, like texts, are readable, with a meaning that is independent of the intentions of the actors and subject to conflicting interpretations. In the same way that a text becomes detached from its author, an action is detached from its agent and may take on unintended meanings of its own. The meaning of an action is then open to an indefinite number of interpretations by an indefinite number of possible readers.

Ricoeur believes that if human action can be read and interpreted like written works, then the methods and practices of textual interpretation can function as a paradigm for the interpretation of action for the social sciences. Ricoeur accepts Wilhelm Dilthey's distinction between two forms of inquiry: scientific explanation of the natural world, and historical understanding of the social world. Yet Ricoeur maintains that hermeneutics is a dialectic of explanation and understanding. Texts and actions have underlying structures to be explained as well as social meanings to be understood.

## METAPHOR AND NARRATIVE

Throughout his career Ricoeur has examined how imaginative and creative uses of language improve our ability to express ourselves and extend our understanding of the world. Symbols, myths, metaphors, and fiction can capture experience in ways that ordinary, descriptive language cannot. Ricoeur maintains that the reference of creative language is divided or split, meaning that such writing points to aspects of the world can only be suggested and referred to indirectly. Creative language refers to such aspects of the world as if they were real and as if we could be there. In *The Rule of Metaphor* (1975),

Ricoeur develops his thesis that the split-reference of creative discourse discloses a possible way of being in the world that remains hidden from ordinary language and first-order reference. A metaphor is a heuristic fiction" that redescribes reality by referring to it in terms of something imaginative or fictitious, allowing us to learn something about reality from fiction. Heuristic fictions help us to perceive new relations and new connections among things, broadening our ability to express ourselves and understand ourselves.

Like all discourse, a metaphor is a communicative utterance that is produced as event, but understood as meaning. Yet, only a live metaphor is at the same time both event and meaning. A dead metaphor has lost its event character when it becomes a commonplace expression, such as, for example, to describe someone who is nervous as having butterflies in their stomach. A live metaphor contains a metaphorical twist that produces a new, surprising meaning. The meaning results from a tension in the way something is described metaphorically and how we normally understand it to be. In order to grasp the differences and resemblance that constitute a metaphor, we must see through the first-order, ostensive reference to the second-order, creative reference to understand how it relates the world. To understand what a metaphor means is to see that it is similar to and different from an ordinary description. The tension in a living metaphor between literal and imaginative must be preserved, not overcome, to be understood. Ricoeur argues that living metaphors create new interpretations that may potentially transform the way we understand and act.

In his three-volume *Time and Narrative* (1983, 1984, and 1985), Ricoeur continues to develop the themes of semantic innovation and the ability of poetic discourse to disclose new ways to see and to be in the world. The basic unit of a narrative is a plot, which unifies the elements of a story, including the reasons, motives, and actions of characters with events, accidents, and circumstances together into a coherent unity. A plot synthesizes, integrates, and schematizes actions, events, and, ultimately, time into a unified whole that says something new and different than the sum of its parts. The thesis of *Time and Narrative* is that connection exists between the temporal character of human experience and the act of narrating a story. Temporal experience is expressed in the form of a narrative, as a narrative is able to reflect our social reality because it expresses temporal experience.

The circularity of time and narrative is mediated by three senses of representation: *mimesis*<sub>1</sub>, *mimesis*<sub>2</sub>, and *mimesis*<sub>3</sub>. *Mimesis*<sub>1</sub>, or prefiguration, represents the aspect

of the imitation of action that draws on our pre-understanding of the difference between human action and physical activity. It has three aspects—structural, symbolic, and temporal—that form the cognitive and practical background that determines how we interpret human action. *Mimesis*<sub>2</sub>, or configuration, is the pivot of the analysis of the relationship between time and narrative in which actions are configured into a story by means of a plot. An action becomes an event in relation to a plot of a story. In turn, a story is more than a succession of actions but rather the organization of events into an intelligible whole. *Mimesis*<sub>3</sub>, or refiguration, refers to the act of reading that changes our practical understanding according to the configuration of the story. The act of telling and interpreting stories links narration with the practical transformation of the world. Hermeneutics for Ricoeur is now construed as the telling, writing, and understanding of fictional and nonfictional stories, in effect, linking time, narratives, and history.

In *Time and Narrative*, Ricoeur introduces the idea of a narrative identity. His thesis is that we understand a person's identity as we would a character in a fictional or historical narrative. One's self-identity is constituted by means of *emplotment*, which configures and synthesizes diverse and multiple elements of a life into a unified whole. Just as the story of a life unfolds like a narrative, the identity of a character also unfolds in a narrative. One's identity is constituted by the stories told about oneself, as well as the stories told by other significant figures in our lives such as parents, spouses, friends, and enemies. A personal identity is also tied to larger group identities, which, similar to a personal identity, are partly chosen, partly inherited, and constituted by the stories we tell about it. The identity of a group, culture, or nation requires that its members are convinced of the truth and rightness of their story. To be effective, these narratives have to shape how the members understand themselves as a part of the group. Ricoeur is particularly interested by stories of founding events that establish and sustain communities, and form our individual and group identities.

## ETHICS AND POLITICS

Ricoeur is also interested in the role narratives play in moral deliberation. The ethical implications for personal or narrative identity is that an agent must maintain some kind of continuity of time in order to be accountable for one's actions. Identity is constitutive of accountability; narrative is constitutive of identity. In addition to attributing actions to agents, narrative discourse also attributes moral obligations to agents who have the

power to act and who are capable of being acted upon. Narration further mediates between description and prescription by providing a context and characters in ethical questioning. We tell stories as a part of the thought experiments we conduct, which allow us to test moral judgments in imaginary cases. Narration thus forms not only our moral ideals but also the stories we tell of ourselves and each other that help us determine if we have achieved it. How we individually and collectively remember events is crucial to the way we hold others accountable for their actions.

The fourth set of studies in *Oneself As Another* (1990), following the studies on speaking, acting, and narrating, form what Ricoeur ironically calls his "little ethics," an ambitious attempt to mediate between an Aristotelian, teleological conception of the ethical aim and a Kantian, deontological conception of the moral norm. Ricoeur's notion of practical wisdom incorporates the idea from the Aristotelian heritage that ethics is the practice of becoming a good person as a member of a political community. The good life consists in developing the virtues, habits, and practices that enable us to develop ourselves, sustain interpersonal relationships, and create a life of happiness together.

Ricoeur also incorporates the idea from the Kantian heritage that morality is defined by the obligation to respect universal moral norms. In this tradition, moral actions must be motivated solely out of duty to the moral law. Morality consists in obeying moral laws that are binding on everyone, respecting the dignity of other people, and acting as an autonomous member of a moral community. For Ricoeur, practical wisdom is the art of mediating the particular requirement of the ethical aim and the universal requirement of the moral norm geared toward acting appropriately and justly in order to achieve happiness with others in a good and just society.

In *Oneself As Another*, Ricoeur proposes three theses with respect to ethics and morality: 1) the primacy of ethics over morality; 2) the necessity that the ethical aim be mediated by the moral norm; 3) morality must seek recourse in ethics to resolve conflicts and aporias. Ethics encompasses morality—but while it is subordinate to ethics, morality is a necessary, deontological moment of the actualization of ethics. The final recourse to ethics (informed by morality) is a form of practical wisdom geared toward the appropriate application of universal norms in particular situations. The reason why ethics needs morality is to ensure that ethical life respects the autonomy and dignity of everyone.

The reason why morality needs ethics is twofold: 1) Without ethics morality would be empty; it is founded on and presupposes our desire to live well together with others; 2) When deontological norms produce conflicting obligation—as they inevitably do—we must refer back to the ethical aim of a particular good life in order to figure out what to do. Sometimes there is no right answer to moral problems. If moral judgment were simply a matter of balancing the ethical aim and moral norm, there would be no room for the tragedy of action, exemplified in stories similar to that of Sophocles's *Antigone*. It is in these intractable situations that the art of practical wisdom helps us make decisions and act justly and appropriately in the face of tragic situations.

Ricoeur's main contribution to political philosophy is his notion of the political paradox. He maintains that, on one hand, political authority is legitimate if it comes from the rational consent of the governed; on the other hand, political practice is often coercive, even violent, which is something, in principle, to which individuals cannot consent. The paradox of political authority is permanent. Ricoeur agrees with Hannah Arendt that it is necessary to distinguish legitimate power-in-common from illegitimate power-over, but he agrees with Max Weber that political institutions are in fact often characterized by domination. Consequently, we should recognize that political power and political discourse always teeters at the edge of violence and illegitimacy. The political sphere is a fragile balance between authority and force, reason and tradition, ideology and utopia.

In *The Just* (2000) Ricoeur argues that coping with political power is an exercise in practical wisdom, a mediation of our desire to live together in communities with the requirement of justice and the rule of law. Social justice not only requires democratic political and economic institutions that respect human rights, treat people equally, protect our liberties, and allow for full political participation, but it should aim to foster a good life for communities, emphasize the membership of citizens for whom political participation matters, and recognize the plurality of social goods and historic values that make us who we are.

Ricoeur's more recent work, *Memory, History, and Forgetting* (2004), examines the role memories play in our ability to represent the past and make present something that is absent. The first part is a meticulous phenomenology of memory, examining the object of memory, the act of remembering, and the nature of personal and collective memories. The second part is an epistemology of history that examines the documentary phase of archiving

eyewitnesses, the explanatory phase where historical explanations occur, and the representative phase where history takes its written or literary form. The third part is a hermeneutics of our historical conditions that examines the limits to our historical knowledge, the existential and temporal conditions of our historical knowledge, and the role of forgetting in relation to memory and history.

The work concludes with a plea for forgiveness as the best way to remember events in order to right past wrongs and to restore social bonds. Individual and groups must learn to remember events differently if they wish to achieve recognition and potentially reconcile. The political implications of memory involve policy considerations for the just allotment of memory to redress excesses of both memory and forgetting. Forgiveness, however, goes beyond justice and approaches the realms of charity and gift-giving. To ask for forgiveness is to recognize that a crime may be unforgivable. Yet, Ricoeur maintains that forgiveness is the best way to remember events to permit more hopeful futures together.

In the 1950s, Ricoeur rivaled Sartre in popularity in France. By the end of the 1960s his popularity waned, along with other phenomenologists, as a new generation of intellectuals dominated the French scene. Ricoeur spent much of the 1970s in the United States, writing in English and assimilating the work of Anglo-American philosophers. His French readers, however, were perplexed by his turn to analytic philosophy and, although his works continued to be read by theologians, his reputation suffered among philosophers. However, with the publication of the highly acclaimed *Time and Narrative* in the 1980s, Ricoeur was again recognized as among France's leading intellectuals. By then he had outlived and surpassed the generation of postmodernist philosophers from the late 1960s, taking his own creative, literary turn without ever abandoning his conviction that philosophy and reason are synonymous. Over his career, Ricoeur received honorary degrees from approximately forty universities, delivered the prestigious Gifford Lectures in 1986, and was awarded numerous significant international prizes, including the 2004 John W. Kluge Prize for Lifetime Achievement in the Human Sciences. After his retirement, he resided outside of Paris and continued to publish well into his nineties.

**See also** Continental Philosophy; Ethics; Hermeneutics; Metaphysics; Philosophy of Language; Philosophy of Religion; Social and Political Philosophy.

**Bibliography**

## PRIMARY WORKS

- Gabriel Marcel and Karl Jaspers. Philosophie du mystère et philosophie du paradoxe.* Paris: Temps Present, 1948.
- Freedom and Nature: The Voluntary and The Involuntary* (1950). Translated by Erazim Kohak. Evanston, IL: Northwestern University Press, 1966.
- History and Truth* (1955). Translated by Charles A. Kelbley. Evanston, IL: Northwestern University press, 1965.
- Fallible Man* (1960). Translated by Walter J. Lowe. New York: Fordham University Press, 1986.
- Freud and Philosophy: An Essay on Interpretation* (1965). Translated by Denis Savage. New Haven, CT: Yale University Press, 1970.
- The Symbolism of Evil* (1960). Translated by Emerson Buchanan. New York: Harper and Row, 1967.
- Husserl: An Analysis of His Phenomenology.* Translated by Edward E. Ballard and Lester G. Embree. Evanston, IL: Northwestern University Press, 1967.
- The Conflict of Interpretations: Essays in Hermeneutics* (1969). Translated by Willis Domingo. Evanston, IL: Northwestern University Press, 1974.
- Political and Social Essays.* Translated by Donald Stewart. Athens: Ohio University Press, 1974.
- The Rule of Metaphor: Multi-Disciplinary Studies in The Creation of Meaning in Language* (1975). Translated by Robert Czerny with Kathleen McLaughlin and John Costello. London: Routledge and Kegan Paul, 1978.
- Interpretation Theory: Discourse and The Surplus of Meaning.* Fort Worth: Texas Christian Press, 1976.
- The Philosophy of Paul Ricoeur: An Anthology of His Work,* edited by Charles E. Reagan and David Stewart. Boston: Beacon Press, 1978.
- Hermeneutics and The Human Sciences: Essays on Language, Action, and Interpretation.* Translated by John B. Thompson. Cambridge, U.K.: Cambridge University Press, 1981.
- Time and Narrative.* 3 vols. (1983, 1984, and 1985). Translated by Kathleen McLaughlin and David Pellauer. Chicago: University of Chicago Press, 1984, 1985, and 1988.
- Lectures on Ideology and Utopia.* Translated by George H. Taylor. New York: Columbia University Press, 1985.
- From Text to Action: Essays in Hermeneutics II* (1986). Translated by Kathleen Blamey and John B. Thompson. Evanston, IL: Northwestern University Press, 1991.
- À l'école de la philosophie.* Paris: J. Vrin, 1986.
- Le mal: Un défi à la philosophie et à la théologie.* Geneva: Labor et Fides, 1986.
- Oneself as Another* (1990). Translated by Kathleen Blamey. Chicago: University of Chicago Press, 1992.
- A Ricoeur Reader: Reflection and Imagination,* edited by Mario J. Valdes. Toronto: University of Toronto Press, 1991.
- Lectures I: Autour du politique.* Paris: Seuil, 1991.
- Lectures II: La contrée des philosophes.* Paris: Seuil, 1992.
- Lectures III: Aux frontières de la philosophie.* Paris: Seuil, 1994.
- The Just* (1995). Translated by David Pellauer. Chicago: University of Chicago Press, 2000.
- Thinking Biblically: Exegetical and Hermeneutical Studies.* Chicago: University of Chicago Press, 1998.

- Critique and Conviction* (1995). Translated by Kathleen Blamey. New York: Columbia University Press, 1998.
- What Makes Us Think? A Neuroscientist and A Philosopher Argue About Ethics, Human Nature, and The Brain.* Translated by M. B. DeBevoise. Princeton, NJ: Princeton University Press, 2000.
- La mémoire, l'histoire, l'oubli.* Paris: Seuil, 2000.
- Le Juste II.* Paris: Esprit, 2001.
- Parcours de la reconnaissance.* Paris: Stock, 2004.
- Sur la traduction.* Paris: Bayard, 2004.

## SECONDARY WORKS

- Abel, Olivier. *Paul Ricoeur: la promesse et la règle.* Paris: Michalon, 1996.
- Bourgeois, Patrick. *Extension of Ricoeur's Hermeneutic.* The Hague: Martinus Nijhoff, 1975.
- Cohen, Richard A. and James L. Marsh, ed. *Ricoeur As Another: The Ethics of Subjectivity.* Albany: State University of New York Press, 2002.
- Dauenhauer, Bernard. *Paul Ricoeur: The Promise and Risk of Politics.* Lanham, MD: Rowman & Littlefield Publishers, 1998.
- Dosse, Francois. *Paul Ricoeur: Les sens l'une vie.* Paris: La Découverte, 1997.
- Greisch, Jean. *Paul Ricoeur: L'intinéraire du sens.* Paris: Jérôme Millon, 2001.
- Hahn, Lewis Edwin, ed. *The Philosophy of Paul Ricoeur.* Chicago: Open Court, 1987.
- Ihde, Don. *Hermeneutic Phenomenology: The Philosophy of Paul Ricoeur.* Evanston, IL: Northwestern University Press, 1971.
- Jervolino, Domenico. *The Cogito and Hermeneutics: The Question of The Subject in Ricoeur.* Dordrecht; Boston: Kluwer Academic Publishers, 1990.
- Kaplan, David M. *Ricoeur's Critical Theory.* Albany: State University of New York Press, 2003.
- Kearney, Richard. *On Paul Ricoeur: The Owl of Minerva.* Burlington, VT: Ashgate Publishers, 2004.
- Klemm, David E. *The Hermeneutical Theory of Paul Ricoeur.* Lewisberg, PA: Bucknell University Press, 1983.
- Leeuwen, T.M. van. *The Surplus of Meaning: Ontology and Eschatology in The Philosophy of Paul Ricoeur.* Amsterdam: Rodopi, 1981.
- Reagan, Charles E. *Paul Ricoeur: His Life and His Work.* Chicago: University of Chicago Press, 1996.
- Schweiker, William ed. *Paul Ricoeur and Contemporary Moral Thought.* New York: Routledge, 2003.
- Thompson, John B. *Critical Hermeneutics: A Study in The Thought of Paul Ricoeur and Jürgen Habermas.* Cambridge, U.K., and New York: Cambridge University Press, 1981.
- Van den Hengel, J. *The Home of Meaning: The Hermeneutics of The Subject of Paul Ricoeur.* Washington, DC: University Press of America, 1982.
- Venema, Henry Isaac. *Identifying Selfhood: Imagination, Narrative, and Hermeneutics in The Thought of Paul Ricoeur.* Albany: State University of New York Press, 2000.
- Wood, David ed. *On Paul Ricoeur: Narrative and Interpretation.* New York: Routledge, 1991.

David M. Kaplan (2005)

## RIEHL, ALOIS

(1844–1924)

Alois Riehl, the Austrian neo-Kantian philosopher, was born in Bolzano. Riehl was consecutively Privatdozent (1870), extraordinary professor (1877), and professor (1878) at the University of Graz. He moved to the University of Freiburg in 1882, to Kiel in 1895, to Halle in 1898, and to Berlin in 1905.

Riehl's first philosophy was a realistic metaphysics based on Johann Friedrich Herbart and indirectly on Gottfried Wilhelm Leibniz, and it is of interest, just as in the case of Immanuel Kant, to study the relation between Riehl's precritical and critical writings. Between 1870 and 1872 Riehl made his first realistic, monistic, evolutionist decisions within that dogmatic framework. His *Realistische Grundzüge* (Graz, 1870) centered on the problem of sensation, which he originally conceived as a polycentric reciprocal matrix of consciousness and movement. In *Über Begriff und Form der Philosophie* (Berlin, 1872) he advocated a critical, rational requirement and the scientific character of philosophy, to which he assigned the historical task of leading to ideal ends. In *Moral und Dogma* (Vienna, 1872) he defended the independence of positive morality from beliefs.

A profound study of Kant freed Riehl from his metaphysical dogmatism. The first volume of his *Der philosophische Kritizismus* (1876) marked an important date in the history of the new Kantianism. This work highlighted the hold on Kant of the spirit of the new positive science (not so much through the influence of René Descartes as through that of John Locke and David Hume). Combating psychological and idealistic "misconceptions" of Kant's views, Riehl proposed that the evolution of Kant's thought be studied, and successive editions of *Der philosophische Kritizismus* benefited from previously unpublished writings of Kant discovered by Kant philologists. Kant, according to Riehl, clarified the method of philosophy; in abandoning metaphysics but not identifying itself with science, philosophy shows itself to be theory of knowledge and the methodology of the natural sciences. It is false, however, to eliminate the thing-in-itself and the presupposition of realism common to the sciences, as Hermann Cohen did. Kant distinguished form from content and sought to determine the formal a priori of nature in general and not the particular laws of nature evident in the real experience of the sciences.

In the second and third volumes of *Der philosophische Kritizismus* (1879 and 1887) Riehl reassessed and

amplified his own views. It was not easy: To do so he had to fight with Kant himself (whom Eugen Dühring had blamed for having "two centers of gravity"), even reduced to the first *Critique* alone. In Riehl's view, neither dogmatic realism nor idealism, whether phenomenalist, or absolute, or positivistic, was adequate. Riehl sought to bring Kant up to date concerning the "sensible and logical foundation of knowledge" by surveying the great scientific innovations since Kant's day, such as Robert Mayer's principle of the conservation of energy and the Darwinian theory of evolution. Only then could Riehl critically resume his own realistic monism centered on perception. But perception, the first cognition, is not, in Riehl's judgment, the first reality. The two aspects of perception—the mechanical, which can be made objective and is quantitatively determinable by positive science, and the qualitative, which is subjectively immediate and the sole revealer of the real universal reciprocity—are both phenomenon (*Erscheinung*), although not merely appearance (*Schein*); neither of the two aspects makes up "nature in itself." The monistic propensity, leading to the threshold of metaphysics, comes upon reefs the critique must steer clear of. For example, he desires that his identification of the physical and the psychical should not be confused with materialism, or monadism, or universal psychophysical correspondence, or Spinozistic panpsychism. Again, although Riehl saw mental life as a product of natural evolution, he denied the evolutionary genesis of logical and mathematical concepts.

In 1883, in his inaugural lecture at Freiburg, "Über wissenschaftliche und nichtwissenschaftliche Philosophie," Riehl turned to other fields of philosophy with a progressive valuation (compare a lecture at Princeton, 1913: "Der Beruf der Philosophie in der Gegenwart"). Even in *Der philosophische Kritizismus*, confined to the naturalistic horizon, he had apologized for glancing at "the field of practical philosophy" but had at the end intimated that beyond the realm of science lay the realms of moral action and artistic production (to which he later added religion). It may be asked whether there could be a philosophy of these things if "theoretical" is identical with "scientific" (*wissenschaftliche*). In his later years Riehl struggled with this problem, surrounded too by the other neo-Kantian movements. "Feeling," which he had acknowledged as another side of experience, might be available for that theoretical purpose, but to be so available its evaluations must be freed from practical empiricism. Heinrich Rickert, who had frequent contact with Riehl, later sought to show Riehl's increasing interest in the world of values, until Riehl finally acknowledged that

the role of philosophy is “to raise to conceptual clarity our knowledge of values and their system.”

*See also* Cohen, Hermann; Descartes, René; Dühring, Eugen Karl; Herbart, Johann Friedrich; Hume, David; Idealism; Kant, Immanuel; Leibniz, Gottfried Wilhelm; Locke, John; Neo-Kantianism; Realism; Rickert, Heinrich; Value and Valuation.

## Bibliography

### WORKS BY RIEHL

*Der philosophische Kritizismus*, 3 vols. Leipzig, 1876–1887; 2nd ed., Leipzig: W. Engelmann, 1908–1926; 3rd ed., Vol. I, Leipzig: Kröner, 1924.

*Friedrich Nietzsche: Der Künstler und der Denker*. Stuttgart, 1897.

*Zur Einführung in die Philosophie der Gegenwart*. Leipzig: Teubner, 1903.

“Helmholtz in seinem Verhältnis zu Kant.” *Kantstudien* 9 (1904): 261–285.

“Logik und Erkenntnistheorie.” *Kultur der Gegenwart* 1 (1907): 73–102.

*Philosophische Studien aus vier Jahrzehnten*. Leipzig, 1925. Essays.

### WORKS ON RIEHL

Hofmann, Paul. “Riehls Kritizismus und die Probleme der Gegenwart.” *Kantstudien* 31 (1926): 330–343.

Jaensch, Erich. “Alois Riehl Der Mann und das Werk.” *Kantstudien* 30 (1925): III–XXXVI.

Maier, Heinrich. “Alois Riehl.” *Kantstudien* 31 (1926): 563–579.

Rickert, Heinrich. “Alois Riehl.” *Logos* 13 (1924–1925): 162–185.

Siegel, Carl. *Alois Riehl: Ein Beitrag zur Geschichte des Neukantianismus. Festschrift der Universität Graz*. Graz: Leuschner and Lubensky, 1932.

Spranger, Eduard. “Alois Riehl.” *Forschungen und Fortschritte* 20 (June 1944): 129–130.

*Mariano Campo (1967)*

*Translated by Robert M. Connolly*

## RIGHTS

Although ancient ethics used the concepts of property and justice, each of which presupposes something similar to the concept of a right, the concept of a right in the modern sense developed only later. The first philosopher to define a moral right was most likely William of Ockham (c. 1285–1347), who noted that *jus* sometimes refers to the power to conform to right reason. Thus, he integrated the legal concept of dominium or property into the moral theory that the law of nature determines right action. Hugo Grotius (1583–1645) adopted the resulting

theory of natural rights—rights conferred by the law of nature—and made it the basis for his theory of international law. Hobbes and Locke used their conceptions of natural rights to explain the grounds and limits of political obligation. Hobbes (1588–1679) conceives of a right as a liberty of action that is the absence of any contrary obligation. Because the law of nature requires only that one seek peace, in a state of nature where there is no peace one has a natural right to do anything one desires. He infers that citizens can have a duty to obey the sovereign only if they give up most of their natural rights.

Locke (1632–1704), however, uses his theory of inalienable natural rights to limit the authority of the sovereign over the members of a society. He conceives of natural rights in the traditional way as powers of acting in conformity with the natural law and assumes that the law of nature also imposes obligations upon others not to prevent one from exercising these rights. The Lockean theory of inalienable and imprescriptible natural rights inspired the American *Declaration of Independence*, the French *Declaration of the Rights of Man and the Citizen*, and the subsequent development of constitutional law in much of Europe and North America.

Jeremy Bentham (1748–1832) subjected the doctrine of natural rights to severe criticism. The notion that there are natural rights—rights not created by human legislation—is conceptual nonsense because a right of one party implies a corresponding obligation of some second party, and an obligation exists only when commanded by some law enforced with coercive sanctions. Granting the existence of natural rights is morally perverse, for one could appeal to natural rights to justify any arbitrary action because without any law to define their content they would be indeterminate. Finally, the doctrine of inalienable and imprescriptible natural rights is politically dangerous because it would justify anarchy, for the individual’s absolute natural right to liberty or the pursuit of happiness could neither be given up by the citizen nor extinguished by coercive legislation. After the 1950s, few moral philosophers adopted a traditional natural rights theory because of skepticism about both the existence of a natural law expressing the will of God or a Cosmic Reason as well as one’s ability to know what it might command or forbid by the natural light of reason.

## JURISTIC THEORIES

Skeptical as one may be of natural rights, one can hardly doubt the existence of legal rights. A central concern of modern jurisprudence has been to explain the nature of these rights. Although juristic theories are primarily the-



ories about the nature of legal rights, they can be extended to moral rights if one conceives of morality in terms of something like the moral law. John Austin held that a legal right is equivalent to a relative legal duty. Legal duties are imposed upon persons by laws that command some act or forbearance and constitute obligations because they are enforced by sanctions in the event of disobedience. Although some duties, such as the duty to pay one's taxes, are absolute in the sense that they are owed to society in general and not to any assignable individual, a relative duty is owed to one or more determinate second parties. Thus, the creditor's contractual right to be repaid by the debtor is simply the debtor's duty of repayment owed to the creditor viewed from the latter's point of view.

John Salmond (1920) developed the view of Rudolph von Jhering that a legal right is a legally protected interest. The object of any right is the thing—not necessarily a material object—in which the right holder has an interest. Not all of one's interests constitute legal rights, however. Only those protected by the law. For example, the object of one's legal right not to be killed is one's life, probably the most fundamental interest of every individual. This interest is protected by laws that prohibit killing and that punish murderers. Paul Vinogradoff (1928) maintained that there are three elements of any legal right. (1) A legal right must be claimed by some individual or state; (2) this claim must be recognized by organized society as justified from the public point of view; and (3), this declaration of right must be enforced by the legal authorities. Hence, a legal right is a legally protected claim. One's right not to be injured, and, if injured, to sue for damages, illustrates this conception of rights. Jhering and Vinogradoff set the stage for the ongoing contest between interest and will theories of rights. The former hold that the essential function of rights is to protect some interest of the right holder; the latter insist that the function of rights is to give the will of the right holder some privileged legal status.

Karl Llewellyn (1962) defined a right as a future judicial remedy. To say that Jones has a property right to his car is simply to predict how the courts will decide any cases concerning that car. For example, if someone steals the car, a court will hold that the vehicle shall be returned to Jones. Or if someone damages that car and Jones sues for damages, a court will decide in favor of Jones. This theory of rights reflects the rule skepticism of American legal realists. They argued that it is a mistake to identify the law with the general principles or rules written in the law books. What these general statements mean is left

open to the interpretation of the courts and may or may not be applied in practice depending upon how judges choose to decide particular cases. Hence, what the law really is consists of the decisions made by the courts.

Scandinavian legal realists are not rule skeptics. In fact, Alf Ross (1957) identified a legal right with a set of legal rules such that any one of a number of facts legally implies all of a variety of legal consequences. For example, if someone has purchased a thing *or* if someone has inherited a thing *or* if someone has earned a thing, then if another person steals that thing it shall be returned to the person who purchased or inherited or earned it, *and* if a second party damages that thing the second party shall compensate the person who purchased or inherited or earned that thing, and so on. What the legal right to ownership really amounts to is simply the preceding set of legal rules. Thus, the language of rights is a convenient technique of summing up a complex set of connections in the law. If legal rights seem to be some mysterious sort of ideal entities, this is merely because the language of rights has an emotive meaning that seems to give rights a magical power. This theory of rights reflects the concern of Scandinavian legal realists to reject the metaphysical idealism prevalent in continental legal philosophy.

Wesley Newcomb Hohfeld (1919) examined the writings of judges and jurists and concluded that they use "a right" indiscriminately to express any of four very different fundamental legal conceptions. This ambiguity suggests invalid legal reasoning because claims, liberties, powers, and immunities are different legal relations. A legal claim of X against Y is logically correlative with a corresponding legal duty of Y to X. For example, X's legal right not to be struck by Y is the logical correlative of Y's legal duty not to strike X. A legal liberty or privilege of X in the face of Y is simply the absence of any corresponding duty of X to Y. Thus to say that X has a legal liberty to phone Y after midnight is to deny that X has a legal duty not to phone Y after midnight. A legal power of X over Y is the ability to change some legal position of Y by some voluntary action of X. For example, X's power to give Y a book is X's ability to confer ownership of that book upon Y by handing it to Y and saying "I hereby give this book to you." A legal immunity of X against Y regarding some legal consequence C is Y's lack of legal power over X regarding C. Thus, X has a legal immunity against Y that Y not extinguish X's ownership of a book by Y's act of saying "I hereby take ownership of the book away from X." Hohfeld argued that in the strict sense, only legal claims are rights because an essential feature of the concept of a right is that rights and duties are logically cor-

relative. Every right implies a duty with a corresponding content, and every duty implies a corresponding right. He admitted that lawyers often speak of multiple sets of claims as a legal right, but insisted that clarity and precision require conceiving of a legal right as a single legal claim of one individual against one second party.

### CONCEPTUAL ANALYSES

Joel Feinberg (1980) agreed with Hohfeld that rights are claims, but denied that rights and duties are logically correlative. Although every right implies some duty, not every duty implies any corresponding right. For example, one's legal duty to obey the orders of a police officer is imposed by the impersonal law and not owed to the officer. Similarly one's moral duty to sacrifice some of one's wealth to assist those in need does not imply any right of this or that needy individual to one's charity. Even when a right does imply some corresponding duty, it is a mistake to reduce the right to that duty. What is distinctive and most valuable about rights is that they put one in a position to claim and to demand—and not merely request or beg—performance of the duty owed to one. What confers this status of claimant upon the right holder is some set of rules—legal rules in the case of legal rights and moral principles for moral rights. Hence, a right is a valid claim, a claim justified by some appropriate set of rules.

H. J. McCloskey (1959, 1979) denied that rights are claims *against* and argued that they are entitlements *to* do, have, enjoy, or have done. Having purchased a car and obtained a driving license, one has a legal right to drive one's car. This is a right to do something—to drive on public thoroughfares; it is not primarily a claim against policemen and magistrates not to interfere. One possesses and exercises rights; one makes claims but does not possess or exercise them. Although one's moral right to life gives rise to duties of others not to kill one, it is primarily a right to live and preserve one's life. A hermit's right to life is the hermit's right to do whatever is necessary to sustain the hermit's life—including killing and eating animals—although there is no one else on or near the isolated island against whom the hermit could possibly claim the right to do so. Admittedly, the creditor's right to be repaid does hold against the debtor, but this is a special sort of right and not typical of rights in general.

H. L. A. Hart (1982) agreed with Hohfeld that “a legal right” is used to refer to four very different legal relations, but did not conclude that this makes the expression ambiguous. He explained what liberty-rights, claim-rights, power-rights, and immunity-rights have in com-

mon. They all consist of one or more bilateral liberties protected by a perimeter of duties. For example, at the center of one's liberty-right to look over one's garden fence at one's neighbor is one's legal liberty either to look over one's fence at one's neighbor or not to do so. This right does not impose upon one's neighbor any logically correlative legal duty to allow herself to be looked at; she is legally permitted to erect a higher fence or hide behind a screen. Still, this bilateral liberty is protected by a number of duties against interference. One's neighbor has legal duties not to climb over one's fence and assault one or to blind one with a chemical spray. What distinguishes one species of rights from another is the kind of bilateral liberties at their center. Thus, central to one's power-right to contract is one's liberty either to exercise one's legal power to accept an offer or to refrain from accepting it. And central to the creditor's claim-right to be repaid are the liberties to cancel or refuse to cancel the debt and, in the event of nonpayment, to sue or refrain from suing for payment. Thus, a legal right is an individual choice respected by the law. Presumably Hart thought of a moral right as an individual choice respected by the rules of morality. His view that moral rights concern the proper distribution of freedom strongly suggests some such theory.

Hart's respected choice theory of rights is a will theory of rights, but it is a mistake to assume that all will theories are option theories. Feinberg's claim theory of rights is also a will theory because he argues that to have a right is to be in a position to make a claim, to demand performance of some corresponding duty. Yet, his theory does not place any bilateral liberty at the center of every right. He even recognizes mandatory legal rights, such as the right to vote in Australia—where voting is a legal duty—so that one does not have any legal liberty to refrain from voting. As one would expect, Hart argued against interest theories of rights. For one thing, to hold that a right consists in an interest protected by a duty reduces rights to duties that benefit some second party. This renders the concept of a right redundant, for one can say everything one needs to say in the language of beneficial duties. But what is distinctive of rights correlative with duties are powers such as the right holder's power to cancel or enforce performance of that duty. Also, when a right benefits some third party, the right holder is not the party whose interest is protected by the law. Thus, when a parent purchases life insurance and names a child as beneficiary, it is the child's interest that is protected by the insurance company's legal duty to pay, but the right holder is the parent rather than the child.

Carl Wellman (1985, 1995) agreed with Hart that what is distinctive and important about rights is the way in which they allocate freedom and control upon the right holder. He defined a right as a complex of Hohfeldian positions that, if respected, confer dominion over some defining core upon the right holder in the face of one or more second parties. For example, at the core of the creditor's legal right to repayment is the creditor's legal claim against the debtor that the debtor repay the contracted amount at or before the due date. But Hohfeld was mistaken in identifying this legal right with a single legal claim. The creditor's claim would not hold against the debtor unless the right holder also had a legal immunity against the debtor's extinguishing the claim merely by saying "I hereby cancel my debt to you." And it also includes additional associated legal positions, such as the power to sue for repayment in the face of a recalcitrant debtor, the power to cancel the debt if one so chooses, and the legal liberties of exercising these powers. Although this is a modified version of Hart's will theory of rights, Wellman did not put a legal liberty—much less a liberty to choose—at the core of every right. The defining core of a legal right can be a legal claim, power, immunity, or even a liability. He also extended his dominion theory of rights to moral rights by arguing that there are moral liberties, claims, powers and immunities analogous to the legal relations Hohfeld identified.

Joseph Raz (1986) defined rights by their role in practical reasoning rather than in terms of Hohfeld's fundamental legal conceptions. It is a mistake to identify rights with interests, even protected interests, because rights serve as intermediate reasons linking interests to duties. At the same time, a right cannot be reduced to some correlative duty because a right is logically prior to any duty, and a single right can imply more than one duty or various duties under varying circumstances. To say that someone has a right is to say that—other things being equal—an aspect of an individual's well-being (one of that individual's interests) is a sufficient reason for holding some other person or persons to be under a duty. However, what makes one's interest of sufficient importance to ground duties need not be merely the value of that interest to oneself. For example, Abel's right to free speech is based on Abel's interest in speaking freely together with the public interest in allowing citizens to speak their minds without unjustified restrictions. Thus, rights are interest-based reasons for duties. This is an interest theory of rights, but more complex than the traditional protected interest theories of Jhering and Salmond.

Rex Martin (1993) rejected Feinberg's view that rights are valid claims. For one thing, not every right implies some corresponding duty as claim-rights do. The logical correlative of the constitutional right to free speech is a disability of Congress to enact statutes limiting speech. More importantly, moral or legal justification is not sufficient to establish a claim as a right. A slave's claim to freedom, no matter how thoroughly justified by moral principles, would be infirm as a right in any society where others could disregard it with impunity. Even if the slave's claim to freedom were justified by the legal rules of that society, it would be merely a nominal right if public officials, including judges, failed or refused to act in accordance with these rules. Real rights—moral rights as well as legal rights—presuppose the social practices of recognition and maintenance. Thus, rights are established ways of acting or being treated—for example, the civil right to the free exercise of one's religion or the moral right to be rescued from imminent danger. Because rights must be established by social practices, no right can exist independently of the institutions of a society. Thus, human rights are best understood as morally justified civil rights.

Judith Jarvis Thomson (1990) doubted that one can base a theory of rights on a definition of "a right" or general description of the nature of rights. Therefore, she used a conceptual analysis different from those previously described. She suggested that to attribute a right might be to talk about permissible and impermissible actions, but in a way which groups them to bring whole clusters of cases to bear on each other. Hence, to learn what the moral or legal significance of having some right is, one must discover the moral or legal consequences for the right holder and others of that right. For example, Jill has a moral right that Jack not break her nose. This implies—other things being equal—at least that Jack ought not to break Jill's nose, that it is morally permissible for Jill to defend herself against any attempt by Jack to break her nose, and that if Jack does break Jill's nose he ought to pay her medical expenses. Thus, Thomson analyzed rights in terms of what the right holder and others may or ought to do. This is an analysis resting upon the judgments of particular cases, not one derived from general principles.

#### POSSIBLE POSSESSORS

A conceptual analysis of rights usually implies something about the necessary conditions for the possession of any right. Thus, Hart's respected choice conception of rights implies that it is idle and misleading to ascribe rights to young children who have not yet developed the ability to

choose. Neil MacCormick (1982) suggested that children's rights are a test case for any theory of rights. He thought it clear that these children do have moral rights, including the rights to be nurtured, cared for, and if possible, loved. Because will theories cannot explain these rights, they must be rejected. In their place, he proposed a protected interest theory of rights. Because even neonates do have interests, this theory can explain how it is possible for them to be right holders.

Feinberg's theory also seems to imply that very young children could not be right holders. He maintained that to have a right is to be in a position to make a claim, to demand something as one's due. Infants seem incapable of claiming in this performative sense. But Feinberg believed that clearly even wee babies do make claims, not in their own persons but through parents or guardians who act as their representatives. These representatives are claiming on their behalf, acting in their interests. Because children do have interests from the day they are born, they are capable of being represented and, therefore, are possible holders of legal and moral rights. Thus, Feinberg combined a will theory of the nature of rights with an interest theory of possible right holders.

Feinberg extended his theory of possible right holders to the more controversial debate about animal rights. Because many non-human animals do have interests in food, shelter, and freedom from pain, human beings can represent them and make claims on their behalf. Hence, animals are also possible possessors of rights. R. G. Frey (1980) challenged the assumption that animals can have interests in the relevant sense. One can take an interest in something only if one can desire or want that thing, and this requires that one believe something about that thing. But because animals lack any language adequate for believing, they cannot possibly have the desires or wants presupposed by the interest theory of possible right holders.

H. J. McCloskey denied that it is the capacity to have interests that makes one a possible right holder because one may, on moral grounds, choose to exercise one's rights contrary to one's interests. The notion of exercising—acting on the basis of—one's rights is central to the concept of a right. And to exercise or refrain from exercising some moral right requires that one make a moral choice. Hence, it is the capacity for moral autonomy, for self-direction, and self-determination, that is required for the possession of rights. Because animals lack this capacity, they are not possible right holders. Robert Elliot (1987) agreed that the capacity to exercise one's rights is necessary for the possession of rights, yet denied that this

requires full moral autonomy—the ability to consider moral reasons and choose on those grounds. A human being could exercise one's moral right to self-defense simply by unreflectively defending oneself against an attacker. An animal could do the same. Hence, animals are capable of acting in a sense robust enough to enable them to possess rights.

Tom Regan (1983) accepted the Kantian view that human beings have moral rights because of their inherent value, but argued that it is arbitrary to restrict inherent value to moral agents. It is being the subject-of-a-life—having a life that goes better or worse for one—that confers inherent and not merely instrumental value upon one. Because at least the higher animals are also subjects-of-a-life, they are also moral right holders.

Are human fetuses capable of possessing human rights, including the right to life? Mary Anne Warren (1973) granted that unborn children are human in the genetic sense of being members of the same biological species as adult human beings, but denied that this is relevant to whether they are members of the moral community of right holders. It is because normal adult human beings are persons that they possess moral rights. The traits that are central to personhood are consciousness, rationality, self-motivated action, the capacity to communicate, and self-awareness. Although it may be uncertain how many of these and in what degree are required for personhood, it is clear that fetuses possess few of these traits in any significant degree. Therefore, a human fetus cannot possess any significant right to life.

Advocates of the right to life often argue that the human fetus has the capacity to develop into an adult person; this potentiality gives it the moral right to life. The standard reply to this argument is that the potentiality to become a person implies only the capacity to acquire rights in the future, not the capacity to possess them before birth. However, Francis C. Wade (1975) argued that if a kernel of seed corn has the capacity to grow into a stalk of corn, this must be because of something in the present nature of that kernel, an active tendency to grow. Similarly, the human fetus's potentiality of full humanity in the morally relevant sense is an active tendency to develop personhood already existing in the fetus. This explains how fetuses can now possess moral rights, including the human right to life.

## MORAL RIGHTS

Although Bentham (1962) rejected the existence of moral rights independent of the law, John Stuart Mill (1969) defended their existence partly as moral grounds for

judging that some law is unjust. A duty is a kind of action that a person may rightfully be compelled to perform and that one may be punished for not performing. Legal duties ought to be enforced with legal sanctions; moral duties are obligations one should be compelled to perform by public opinion or the internal sanction of one's conscience. A moral right is the logical correlative of a relative moral duty, a duty owed to the right holder because that is the one who would be harmed by its nonperformance. Why ought society to defend one in the possession of one's moral rights? Mill's answer is that this will promote the general utility—the greatest well-being—of all the members of the society.

Ronald Dworkin (1977) argued that no utilitarian theory of rights can take moral rights seriously. Sometimes to say that someone has a right to do something is to say merely that to do so would not be to act wrongly, but to say that someone has a right in the strong sense is to assert that it would be wrong to prevent one from so acting. Thus, one may say that Jones has a moral right to spend his money gambling, although he ought to spend it in a more worthwhile way. The moral rights of the citizen against the state—such as the rights to free speech or to freedom from unreasonable searches—are worth taking seriously only if they are rights in the strong sense. If they were grounded on utility, the government would be justified in infringing them whenever it would be useful to do so. But this would undermine their moral purpose: to give the individual special protection against political interference. Therefore, a theory can take moral rights seriously only if it grounds them either on the human dignity of the individual person or on the ideal of political equality. Thus, the most fundamental moral right must be the right to equal concern and respect.

David Lyons (1994) responded that a utilitarian can take moral rights seriously. Dworkin failed to notice Mill's distinction between expediency and morality. An expedient act—one that has the best consequences—can be morally wrong because it violates a moral obligation. What makes an act a moral obligation is not its utility, but that society would be justified in imposing sanctions—either the disapproval of others or of one's own conscience—upon agents who fail or refuse to act. And enforcing moral obligations is justified by its social utility. For someone to have a moral right to something is for others to have a corresponding moral obligation at least not to injure and perhaps to promote the right holder's interest in that thing. This sort of indirect grounding of moral rights on the utility of the enforcement of the cor-

relative duties does not imply that a right may permissibly be violated whenever it would be expedient to do so.

L. W. Sumner (1987) agreed with Bentham that there are no natural rights, but did not infer that there are no moral rights. Although he rejected Hart's will theory of rights and held that the function of rights is to protect some interest of the right holder, he adopted Hart's view that rights presuppose social practice rules. Legal rules are made and upheld by the officials in some legal system; the rules of the moral code of a society are constituted by the practices of its members. But not every conventional right has moral force. A moral right is a morally justified conventional right, either an existing conventional right that it would be morally justified to retain or one that it would be justified to introduce into the conventional morality. And what justifies moral rights is the valuable consequences of maintaining the social practice rules that confer them upon moral agents. Although moral rights are grounded upon their contribution to human welfare, they can be taken seriously because the rules that confer them often constrain the direct pursuit of social utility.

Jeffrie G. Murphy (1977) argued that there are two very different kinds of moral rights. Autonomy rights mark out the special kind of treatment required to respect the dignity of autonomous rational persons. As Kant recognized, persons are ends—and not means only—and ought not to be sacrificed or used without their consent as instruments or resources for achieving the ends of others. Autonomy rights are grounded on the inherent moral value of autonomous rational agents. But Mill recognized a different function of moral rights: to pick out those moral claims that ought to be protected by society, especially by the law. However, Murphy rejected Mill's utilitarian justification of moral rights and argued that they are grounded on a hypothetical social contract of the sort described by John Rawls (1971). Thus, an individual has a social contract right to X only if a law guaranteeing X to the individual would be unanimously chosen by rational agents who are not aware of what their special circumstances would be in their society.

## HUMAN RIGHTS

Although jurists usually identify human rights with the universal human rights recognized in international law, philosophers tend to view human rights as fundamental moral rights one possesses by virtue of being human. It is this latter sort of view that is most relevant here. Gregory Vlastos (1962) defined a just act as one prescribed exclusively by regard for the rights of all those it affects substantially. Although it is often just to distribute goods

unequally according to the merit of the recipients, equalitarian justice respects the equal human rights of everyone affected. Human rights are necessarily equal because they are grounded on the equal human worth of all persons, however different their individual merits. And what gives all human beings equal worth is the equal intrinsic value of their well-being and freedom. Hence, there are two classes of human rights: rights to goods required for human well-being and rights to fundamental human freedoms.

Alan Gewirth (1982) also grounded human rights upon freedom and well-being, but by a different argument. Human rights are primarily moral claim-rights of individual human beings that entail correlative moral duties of other individuals and organizations. All human beings are actual or potential agents, and human action consists in the voluntary pursuit of goals one values. Anyone who engages in action must presuppose that one has a right to the necessary conditions of prospective purposive action. And one cannot claim this and without self-contradiction deny that all other agents have the same rights. The two necessary conditions of human action are freedom and well-being. Hence, the various human rights to basic freedoms and goods are implied by the necessary presuppositions of human action.

James Griffin (2001) argued that the best account of human rights is one that preserves—but goes beyond the insights of—the traditional natural rights theories. It conceives of human rights as protections of personhood. Personhood should be understood in terms of the various strands of agency. These are autonomy (or making one's own decisions), forming a conception of the good life and being able to pursue it, and freedom from interference from others. Hence, there are three classes of human rights: autonomy rights, welfare rights, and liberty rights. The abstract human rights grounded on personhood are made more determinate by practicalities that spell out what is necessary, given the circumstances, to protect personhood. Thus, there are two grounds of human rights: personhood and practicalities.

Contemporary human rights documents reaffirm the traditional civil and political rights, such as the rights to life, free speech, and a fair trial. Yet they also assert social and economic rights, such as the rights to work, social security, and an adequate standard of living. Maurice Cranston (1967) argued that these supposed welfare rights are not genuine human rights. If they were universal human rights, they would impose upon every society the duties to provide employment, old-age pensions, and all necessary food and medical care for all their citizens.

But many societies lack the resources to provide such welfare benefits to all, and there can be no moral duty to do the impossible. Hence, there can be no human rights that would imply such duties. Civil and political rights, however, require only the appropriate legislation. James W. Nickel (1987) replied that civil and political rights also face the problem of scarce resources. No society can afford a police force adequate to secure the right to life of every citizen and the right to a fair trial can be real only where there is an expensive system of courts and adequate legal assistance for all. Moreover, there are ways to realize human welfare rights in societies with varying levels of affluence. A society can introduce programs to achieve progressively full employment or adequate medical care, prune a welfare right to achieve what is most important in it, or if necessary sacrifice more costly human rights in order to secure those that are affordable.

Cranston also insisted that a human right is something of which no one may be deprived without a grave affront to justice, but that it is not a grave injustice for human beings to lack old-age pensions or the medical care they need. Hence, social security and an adequate standard of living may be moral ideals, but they are not genuine human rights. Robert Nozick (1974) went even further and argued that there can be no basic welfare rights because their implementation would violate justice. Welfare rights would require programs such as Aid to Families with Dependent Children and Medicaid. But to fund these or similar programs, the state would have to tax the affluent and redistribute their wealth to the poor. This would be unjust because it would violate the human right to property of those whose wealth is taken from them without their consent. Nozick bases his argument on his entitlement theory of justice. John Rawls (1971), however, advanced a different theory of justice. According to Rawls, the morally justified principles of social justice are those that rational persons would unanimously choose were they in an original position of equality and unaware of their particular circumstances. He argued that one of these principles is that social and economic inequalities are to be arranged so that they are both (a) reasonably expected to be to everyone's advantage, and (b) attached to positions and offices open to all. This would require some redistribution of wealth to overcome great economic inequalities in a society.

Nozick also proposed a concept of rights that seems to exclude rights to welfare benefits. A moral right is a side-constraint on the pursuit of individual or social goals. For example, the right to property makes it morally wrong for one person to become wealthy by stealing from

anyone who owns something one desires, and the right to liberty implies that it would be morally wrong for the state to suppress opposition by imprisoning its critics. Moral rights reflect the Kantian idea that individuals are inviolable because they are ends in themselves and ought not to be used to achieve the goals of others. On this view, moral rights are negative; they constrain the actions of others by imposing only negative duties not to mistreat right holders in morally impermissible ways such as injuring them or interfering with their freedom of action. But any imagined welfare right would impose positive duties to provide welfare benefits such as old-age pensions or payments for medical care to individual right holders. There can be no human rights of this sort if fundamental moral rights are side-constraints that impose only negative duties.

Henry Shue (1980) rejected the view that the traditional civil and political rights are purely negative and that social and economic rights are positive. Every basic moral right imposes three sets of duties: duties to forbear from depriving right holders of the substance of their right, duties to protect right holders against the deprivation of the substance of their right, and duties to aid right holders in obtaining or regaining the substance of any right of which they have been deprived. For example, the basic right to liberty implies that the state has the duties not to imprison innocent persons, to protect individual persons from being kidnapped, and to aid anyone who has been unjustly imprisoned or kidnapped to obtain her release. Because some of these duties are negative and others positive, it is a mistake to argue that there cannot be any welfare rights, such as the right to subsistence benefits, simply because this would impose positive duties to aid those who lack the means of subsistence. As one would expect, the philosophy of rights remains as controversial today as it has been during the past few centuries.

**See also** Social and Political Philosophy.

### Bibliography

- Austin, John. *Lectures on Jurisprudence, or the Philosophy of Positive Law*. 5th edition, edited by Robert Campbell. London: John Murray, 1885.
- Bentham, Jeremy. *Anarchical Fallacies*. In *Works of Jeremy Bentham*. vol. 2, edited by John Bowring. New York: Russell & Russell, 1962.
- Cranston, Maurice. "Human Rights, Real and Supposed." In *Political Theory and the Rights of Man*, edited by D. D. Raphael. Bloomington, IN: Indiana University Press, 1967.
- Dworkin, Ronald. *Taking Rights Seriously*. Cambridge, MA: Harvard University Press, 1977.
- Elliot, Robert. "Moral Autonomy, Self-Determination and Animal Rights." *The Monist* 70 (1) (1987): 83–97.
- Feinberg, Joel. *Rights, Justice, and the Bounds of Liberty*. Princeton, NJ: Princeton University Press, 1980.
- Finnis, John. *Natural Law and Natural Rights*. Oxford: Clarendon Press, 1980.
- Frey, R. G. *Interests and Rights*. Oxford: Clarendon Press, 1980.
- Gewirth, Alan. *Human Rights*. Chicago: University of Chicago Press, 1982.
- Glendon, Mary Ann. *Rights Talk*. New York: Free Press, 1991.
- Griffin, James. "First Steps in an Account of Human Rights." *European Journal of Philosophy* 9 (3) (2001): 306–327.
- Grotius, Hugo. *The Rights of War and Peace*. Translated by A. C. Campbell. New York: M. W. Dunne, 1901.
- Hart, H. L. A. *Essays on Bentham*. Oxford: Clarendon Press, 1982.
- Hobbes, Thomas. *Leviathan*, edited by C. B. Macpherson. Hammondsworth, U.K.: Penguin, 1968.
- Hohfeld, Wesley Newcomb. *Fundamental Legal Conceptions*. New Haven, CT: Yale University Press, 1919.
- Jhering, Rudolph von. *Geist des römischen Rechts*. Leipzig, Germany: Breitkopf und Hartel, 1877.
- Llewellyn, Karl. *Jurisprudence*. Chicago: University of Chicago Press, 1962.
- Locke, John. *Two Treatises of Government*, edited by Peter Laslett. London: Cambridge University Press, 1967.
- Lomasky, Loren E. *Persons, Rights, and the Moral Community*. New York: Oxford University Press, 1987.
- Lyons, David. *Rights, Welfare, and Mill's Moral Theory*. New York: Oxford University Press, 1994.
- MacCormick, Neil. *Legal Rights and Social Democracy*. Oxford: Clarendon Press, 1982.
- Martin, Rex. *A System of Rights*. Oxford: Clarendon Press, 1993.
- McCloskey, H. J. "Moral Rights and Animals." *Inquiry* 22 (1–2) (1979): 23–54.
- McCloskey, H. J. "Rights." *Philosophical Quarterly* 15 (2) (1959): 115–127.
- Melden, A. I. *Rights and Persons*. Oxford: Basil Blackwell, 1977.
- Mill, John Stuart. *Utilitarianism*. In *Collected Works of John Stuart Mill*, vol 10, edited by J. M. Robson. Toronto: University of Toronto Press, 1969.
- Murphy, Jeffrie G. "Rights and Borderline Cases." *Arizona Law Review* 19 (1) (1977): 228–241.
- Nickel, James W. *Making Sense of Human Rights*. Berkeley, CA: University of California Press, 1987.
- Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books, 1974.
- Ockham, William of. *Opus nonaginta Dierum*. In *Opera politica*, vol. 2, edited by R. F. Bennett. Manchester, U.K.: Manchester University Press, 1963.
- Rawls, John. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1971.
- Raz, Joseph. *The Morality of Freedom*. Oxford: Clarendon Press, 1986.
- Regan, Tom. *The Case for Animal Rights*. Berkeley, CA: University of California Press, 1983.
- Ross, Alf. "Tü-Tü." *Scandinavian Studies in Law* 1 (1957): 139–153.

- Salmond, John William. *Jurisprudence*, 6th edition. London: Sweet and Maxwell, 1920.
- Shue, Henry. *Basic Rights*. Princeton, NJ: Princeton University Press, 1980.
- Sistare, Christine, Larry May, and Leslie Francis, eds. *Groups and Group Rights*. Lawrence, KS: University Press of Kansas, 2001.
- Steiner, Hillel. *An Essay on Rights*. Oxford: Blackwell, 1994.
- Sterba, James P. "The Welfare Rights of Distant People and Future Generations: Moral Side-Constraints on Social Policy." *Social Theory and Practice* 7 (1) (1981): 99–119.
- Sumner, L. W. *The Moral Foundation of Rights*. Oxford: Clarendon Press, 1987.
- Thomson, Judith Jarvis. *The Realm of Rights*. Cambridge, MA: Harvard University Press, 1990.
- Tuck, Richard. *Natural Rights Theories*. Cambridge, U.K.: Cambridge University Press, 1979.
- Vinogradoff, Paul. *Jurisprudence*. In *Collected Papers of Paul Vinogradoff*, vol 2. Oxford: Clarendon Press, 1928.
- Vlastos, Gregory. "Justice and Equality." In *Social Justice*, edited by Richard B. Brandt. Englewood Cliffs, NJ: Prentice-Hall, 1962.
- Wade, Francis C. "Potentiality in the Abortion Discussion." *Review of Metaphysics* 29 (2) (1975): 239–255.
- Waldron, Jeremy. *Liberal Rights*. Cambridge, U.K.: Cambridge University Press, 1993.
- Warren, Mary Anne. "On the Moral and Legal Status of Abortion." *The Monist* 57 (1) (1973): 43–61.
- Wellman, Carl. *Real Rights*. New York: Oxford University Press, 1995.
- Wellman, Carl, ed. *Rights and Duties*, 6 vols. New York: Routledge, 2002.
- Wellman, Carl. *A Theory of Rights*. Totowa, NJ: Rowman and Allenheld, 1985.

*Carl Wellman (2005)*

## RIGNANO, EUGENIO (1870–1931)

Eugenio Rignano was an Italian positivist philosopher and founder (1907) and lifelong editor of the scientific journal *Scientia*. Rignano's first works were sociologically oriented, but he later turned to biology and philosophical biology. His major work, *Psicologia del ragionamento* (1920), places the activity of memory at the basis of all biological and psychic phenomena. Memory is an activity that, through the specific accumulation of concepts, makes possible the progressive adaptation of the organism to the environment, the formation of instincts and emotions, and, in higher organisms, of reasoning. According to Rignano, reasoning is "a series of operations or experiences merely thought out simply"; in other words, a series of operations performed in imagination. The results of these operations are also imagined and are

assumed as the conclusions of the reasoning itself. This conception of reasoning, which Rignano derived chiefly from Ernst Mach, was later applied by him to explain the various kinds of reasoning: intuition, reduction, mathematical and mathematico-logical reasoning, intentional reasoning (dialectical or metaphysical reasoning), and pathological forms of reasoning as well. Rignano stressed the distinction between constructive and intentional reasoning. Constructive reasoning is motivated by a desire to discover the truth, and intentional reasoning by a desire to confirm a truth that a person believes he already possesses. Both types of reasoning utilize the same syllogistic form, but constructive reasoning is characteristic of the positivist scientist and intentional reasoning of the metaphysician. Rignano did not distinguish clearly between logical and psychological considerations; rather, he assumed the psychological mechanism as the basis of the logical validity of reasoning processes. The result is that Rignano's account is not very convincing either as logic or as psychology.

Despite his distaste for metaphysics, Rignano in subsequent works elaborated a kind of biological metaphysics based on the hypothesis that at the foundation of life and its evolution there is a "nervous energy" able to mold organic matter and direct it toward an increasing development and a growing adaptation to the environment. According to Rignano, life in its entirety shows a finalistic aspect that would be inexplicable if it were the product of physicochemical forces. This finalism can be explained, however, by assuming that life is a product of psychic, mnemonic energy, which on the basis of past experience envisions ends of future experience and adapts organic material to those ends. It is a kind of vitalism or animism that, according to Rignano, guarantees to evolution a progressive significance. The progress of evolution continues beyond organic life into moral life. The purpose of moral life is to guarantee to all individuals the satisfaction of their needs and to coordinate these needs in harmonious forms that gradually eliminate conflicts.

**See also** Cybernetics; Italian Philosophy; Mach, Ernst; Memory; Metaphysics; Philosophy of Biology; Positivism; Vitalism.

### *Bibliography*

#### WORKS BY RIGNANO

- Di un socialismo in accordo con la dottrina economica liberale*.  
Turin: Fratelli Bocca, 1901.
- La sociologica nel corso di filosofia positiva di Augusto Comte*.  
Palermo, 1904.



- Psicologia del ragionamento*. Bologna, 1920. Translated by W. A. Holl as *Psychology of Reasoning*. London: K. Paul, Trench, Trubner, 1923.
- La vita nel suo aspetto finalistico*. Bologna, 1922.
- La memoria biologica*. Bologna, 1922. Translated by E. W. MacBride as *Biological Memory*. London: K. Paul, Trench, Trubner, 1926.
- Che cosa è la vita? Nuovi saggi de sintesi biologica*. Bologna, 1926. Translated by N. Mallinson as *The Nature of Life*. London, 1930.
- Man Not a Machine*. London: K. Paul, Trench, Trubner, 1926.
- Problemi della psiche*. Bologna, 1928.

**Nicola Abbagnano (1967)**  
Translated by Nicholas Spadaccini

## RILKE, RAINER MARIA (RENÉ) (1875–1926)

The German poet Rainer Maria (René) Rilke was born in Prague, the son of a minor railway official. His mother, who was of upper-middle-class origin, encouraged him in his early ambition to become a poet. The years 1886–1891, which Rilke spent at military academies in Moravia and Austria, had a traumatic effect on him, and not until 1920 was he able to come to terms with his unhappy childhood and family background. His first volume of poetry, *Leben und Lieder*, appeared in Prague in 1895. Desultory studies, mainly in the history of art, at the universities of Prague, Munich, and Berlin were followed by two journeys to Russia in 1899 and 1900 in the company of Lou Andreas-Salomé, a German-Russian to whom Friedrich Nietzsche had proposed marriage and who later became a follower and friend of Sigmund Freud. During the second of these journeys he met Lev Tolstoy. On his return Rilke joined an art colony in Worswede near Bremen, and early in 1901 he married the sculptress Clara Westhoff, one of its members. They had a daughter, but the short-lived marriage was only an interlude in Rilke's essentially solitary and unsettled life. For the next few years, Rilke's attention was centered on Paris and on Auguste Rodin, to whose work he devoted a monograph in 1903. Although his job as Rodin's private secretary ended in a quarrel, Rilke never ceased to acknowledge the very direct inspiration he received from close daily contact with the sculptor. The first collection of poems that bears the authentic stamp of greatness, *Neue Gedichte I* (Leipzig, 1907), represents Rilke's aim to render in words the immediacy, concreteness, and intensity ("the inward reality") that he discerned in Rodin's work.

With a single-mindedness that has rarely been paralleled in modern literature, Rilke devoted his whole existence to the poetic task he felt called upon to accomplish, subordinating to it all personal and public considerations. The long list of his patrons, most of whom belonged to the aristocracy of central Europe and a few to the German and Swiss patrician bourgeoisie, testifies to the restlessness of his life, and so do his journeys to Sweden (in 1904, on the invitation of Ellen Key), Italy, north Africa (1910–1911), Spain (1913), and repeatedly to France. The long list of his friends (mainly female) and correspondents, among them Paul Valéry and André Gide, includes surprisingly few German writers. Two places were of major importance for the fruition of his poetry: Duino (1910 and 1912), a castle on the Adriatic that belonged to the Princess Marie von Thurn und Taxis-Hohenlohe, where the first Duino Elegy was written, and the little castle of Muzot in the Swiss canton of Valais. It was at Muzot, in February 1922, as the guest of Werner Reinhart, that Rilke, in a storm of inspiration, wrote most of the fifty-five *Sonette an Orpheus* and several smaller collections of poems; and it was there, above all, that he completed his greatest work, which had been interrupted by World War I—the cycle of ten *Duineser Elegien*, several of which were written in the span of a few days. Rilke died at Valmont, Switzerland, after a protracted and painful illness that was diagnosed as leukemia.

Rilke's mature poetry, written after 1907, displays a consistency of attitude and a coherence of poetic devices that make it representative of a whole era of European thought. Following in the wake of Nietzsche, this poetry is informed by an acute historical consciousness. We live in an age when a "religion" that is based on separating transcendence from immanence is no longer viable:

All of the living  
Make the mistake of drawing too sharp distinctions.  
Angels (it is said) would be often unable to tell  
Whether they moved among living or dead.

(FIRST ELEGY)

Our impoverished state is marked by our awareness that "we are not very reliably at home in the interpreted world." Hence, in order to regain for ourselves something that would equal the spiritual and existential fervor that characterized the ages of faith, we must take upon ourselves the task of endowing the world (which, for Rilke, is the world of things and of intimate personal relations) with the inwardness of feeling that other ages directed toward a divinity. Joy, love, and above all suffering and

pain should not be diffuse sensations accompanying an unending series of vague hopes and regrets; they must become the objects of a total commitment. Thus, since we are “not yet” strong enough to give ourselves totally in love, we had better follow the example of the lover (“Gaspara Stampa”) who drew her strength from an unrequited, “uninterrupted” feeling, or indeed of Narcissus, who used the natural world as a magnifying mirror of his feeling. *Les saltimbanques*, the traveling *artistes* of Pablo Picasso’s “blue period” paintings, celebrated in the Fifth Elegy, most fully symbolize our condition. In a world in which all actions are liable to remain uncompleted (“We, though, while we’re intent upon one thing, / can feel the cost and conquest of the other”), suffering—the fullest possible realization and appropriation by the self of what is inflicted from without—will be the greater virtue:

Killing merely is a form of our wandering sadness ...

Pure in the spirit serene

Is what we ourselves endure.

History, for Rilke, is not a social phenomenon but a pageant of situations and persons in whom the ideal of completion and strength of feeling was realized, just as the contemporary world is a series of images that portray our deprivations and stunted responses. To make something of one’s fate, of one’s experiences, is to give them the permanence (essentially poetic) of a moment of intensity. Similarly, the supreme task, set by the imminence of death, is to repair the adventitiousness of death by drawing it into my life, by making of it “my own death.”

The immensity of the task of creating a new spirituality is betrayed by the complex, and quite conscious, ambiguities of Rilke’s images of transcendence, chief of which is the image of the Angel, as he appears in the Elegies. He is a messenger (*angelos*) from another sphere; hence, there must be one who sent him. But the Angel comes upon us with a terrible majesty and strength which, to us who are weak, is all his own. In many such astonishing images Rilke expresses the “pure [=necessary] contradiction” that he sees as the root of our being: only by living in total commitment to “the Earth,” the here and now, can man transform it into “the heart’s inner space,” and thus wrest some eventual transition into a “soundless” Beyond—wrest it from he knows not whom. The most accomplished practitioner of such transformations is Orpheus, the poet-maker who, in the creative act, stills all strife by transforming it into song, eternalizes the moment by making of it a monument of inwardness, and

transfixes suffering into the eternally valid image of “Lament” (Tenth Elegy).

In a world yearning for the security of faith and finding it in ideology, Rilke’s vertiginous images were reduced to prosy precepts for living, becoming thus at once esoteric and banal. Rilke’s poetry is not necessarily esoteric, and the creative activity he extolled is closely related to the poetic; but he addressed himself to the single individual. The social sphere of modern life is branded as wholly inauthentic (Rilke either ignored or briefly satirized it); all concerted action is an escape from defective selfhood. He understood and expressed velleities supremely well; his poetry hardly offers a nostrum to cure them.

*See also* Freud, Sigmund; “My Death”; Nietzsche, Friedrich; Valéry, Paul.

### **Bibliography**

#### WORKS BY RILKE

Only the most recent four-volume edition, *Gesammelte Werke*, edited by Ernst Zinn (Frankfurt, 1955–1961), shows the magnitude of Rilke’s work. In addition, some eight collections of his letters are of major importance. The two-volume edition of *Selected Works* (London, 1954–1960) includes J. B. Leishman’s brilliant translations of all of Rilke’s major poetry.

#### WORKS ON RILKE

No English biography has appeared since E. M. Butler, *Rainer Maria Rilke* (Cambridge, U.K.: Cambridge University Press, 1941). H. E. Holthusen, *Rainer Maria Rilke in Selbstzeugnissen und Bilddokumenten* (Hamburg: Rowohlt, 1958), is a succinct and discriminating account of the poet’s life. See also N. Purtscher-Wydenbruck, *Rilke: Man and Poet* (London, 1949), and E. Buddeberg, *Rainer Maria Rilke—Eine innere Biographie* (Stuttgart: Metzler, 1955). A brilliant assessment of Nietzsche’s influence on Rilke is Chapter 5 of Erich Heller, *The Disinherited Mind* (Cambridge, U.K.: Bowes & Bowes, 1952). E. C. Mason has devoted many studies to a critique of Rilke’s work, among them *Lebenshaltung und Symbolik bei R. M. Rilke* (Weimar, Germany: H. Böhlau Nachf, 1939) and *Der Zopf des Münchhausen* (Einsiedeln, Germany, 1949). Two exegetic works of major importance are Jacob Steiner, *Rilke’s Duineser Elegien* (Bern: Francke, 1962), and Romano Guardini, *Rilke’s Duino Elegies: An Interpretation*, translated by K. G. Knight (London, 1961).

**J. P. Stern (1967)**

## RINTELEN, FRITZ- JOACHIM VON (1898–1979)

Fritz-Joachim von Rintelen, the German philosopher of value, was born in Stettin. He received a doctorate in philosophy in 1924 from the University of Munich, where he began to lecture in 1928. Von Rintelen was appointed professor at the University of Bonn in 1933 and at Munich in 1936, but he was suspended on political grounds in 1941. In 1947 von Rintelen became professor of philosophy, psychology, and pedagogy at the University of Mainz. He was a visiting professor at the Universidad Nacional de Córdoba, Argentina, from 1951 to 1952 and at the University of Southern California in 1957.

Von Rintelen gave both a systematic and a detailed historical interpretation of the problem of value. Prior to his dismissal from his professorship at Munich, he had explicated in detail a theory of value (*Wert*) and of meaning (*Sinn*) and had built a philosophical anthropology upon it. After World War II he applied this theory to an analysis and penetrating criticism of the irrationalistic, nihilistic, and pessimistic currents in contemporary European philosophy and literature, showing how the theory resolves the conflicts and paradoxes that he reveals in these currents.

His doctrine of values and of personality is rooted in the realistic tradition of Platonism and Scholasticism but also shows the influence of German idealism. The chief direction of his thought was set in his two academic dissertations, a criticism of the pessimistic philosophy of religion of Eduard von Hartmann and an attempt to extend Ernst Troeltsch's efforts, in the later years of his life, to overcome historical relativism through a theory of values and their operation in history.

Two points in von Rintelen's criticism are particularly salient. The first is his attack upon all dualisms of intellect and will or of mind and life (Max Scheler), all subordination of the rational to a more inclusive irrational, and every combination of an idealistic theory of scientific knowledge with a realism in metaphysics. To these distinctions he opposes an ontological interpretation of value by which these dualistic tensions can be resolved. He rejects von Hartmann's teleology of self-destruction as an ontological impossibility and an aesthetic misreading of the tragedy of our culture; this tragedy cannot be denied, but it implies a transcendent normative meaning to be attained through the reflective transformation of our actions. Thus, there is an inclusive ontological meaning, not attainable through scientific

logic but through the value experiences of life, which sustains the human spirit and human life.

Von Rintelen's ontological theory of real value (*Realwert*) was constructed in opposition to psychological, positivistic, and phenomenological definitions. Real value is an objective context of meaning that can be particularized and made concrete through conscious or unconscious strivings. Each actualized value possesses an intrinsic worth varying in intensity with its degree of meaning and a relational worth by virtue of which it enters into a wider order of values. Thus every real value is vertically capable of degrees of normative validity and historically capable of individualization within larger contexts of culture and of personal action. Values are individualized in two spheres corresponding to life and mind in man. In the sphere of nature, objects are primarily existent and only secondarily valued; in that of mind or personality, objects are primarily mental and only secondarily grounded in concrete existence. From this viewpoint human history can be understood as a continually renewed effort to actualize values in terms of a personal regulative ideal of the highest possible fulfillment and in relation to an ultimate *summum bonum*, God, the unique, autonomous, and inclusive real value.

In 1932 von Rintelen published the first volume (ancient and medieval) of a historical study in which the development of this theory of value was to be traced in European thought. This work was left incomplete but was supplemented by specialized historical and systematic articles.

In his later critiques of existentialism and other contemporary intellectual currents, von Rintelen analyzed the plight of man as portrayed in modern philosophy and literature and, by correcting the subjectivism and finitism implicit in this portrayal through his own doctrine of value transcendence, points out the way to "a rewon security of spirit." Outstanding among these works are *Philosophie der Endlichkeit* (1951), which includes analyses of Martin Heidegger, Rainer Maria Rilke, Gabriel Marcel, and Jean-Paul Sartre, and *Der Rang des Geistes* (1955), a thorough and distinguished study of Johann Wolfgang von Goethe as philosopher, in which the inner tensions or polarities of the poet's thought are examined and Goethe's movement from an eclecticism to a rationally justified theism and an operative human ideal of rational freedom and love is portrayed. In these books von Rintelen shows himself not merely as a constructive philosopher but also as an able critic of literature and culture.

Von Rintelen's thought may thus be considered as a reconstruction of the Christian intellectual tradition in which the inevitable tragedy that inheres in the polarities of human existence may be overcome through a transcendent order of values in which meaning and impulse are harmonized.

**See also** Existentialism; Goethe, Johann Wolfgang von; Hartmann, Eduard von; Heidegger, Martin; Idealism; Intrinsic Value; Marcel, Gabriel; Philosophical Anthropology; Platonism and the Platonic Tradition; Rilke, Rainer Maria (René); Sartre, Jean-Paul; Scheler, Max; Troeltsch, Ernst; Value and Valuation.

### **Bibliography**

Works by von Rintelen include *Pessimistische Religionsphilosophie der Gegenwart: Untersuchung zur religionsphilosophischen Problemstellung bei Eduard von Hartmann und ihre erkenntnistheo-retischmetaphysischen Grundlagen* (Munich, 1924); "Der Versuch einer Überwindung des Historismus bei Ernst Troeltsch," in *Vierteljahresschrift für Literaturwissenschaft und Geistesgeschichte* 8 (1930): 324–372; *Der Wertgedanke in der europäischen Geistesentwicklung*, Part I, *Altertum und Mittelalter* (Halle, 1932); *Dämonie des Willens* (Wiesbaden, 1948); *Von Dionysos zu Appollon. Der Aufstieg im Geiste* (Wiesbaden, 1948); *Philosophie der Endlichkeit als Spiegel der Gegenwart* (Meisenheim am Glan, Germany: Hain, 1951), partly translated by Hilda Graf as *Beyond Existentialism* (London: Allen and Unwin, 1963); and *Der Rang des Geistes. Goethes Weltverständnis* (Tübingen, 1955).

Other works by von Rintelen are *Der Aufstieg im Geiste*, 2nd ed. (Frankfurt: Metopen-Verlag, 1968); *Johann Wolfgang von Goethe: Sinnerfahrung und Daseinsdeutung* (Munich: E. Reinhardt, 1968); *Contemporary German Philosophy and Its Background* (Bonn: Bouvier, 1970); *Values in European Thought* (Pamplona: Universidad de Navarra, 1972–); *Philosophie des lebendigen Geistes in der Krise der Gegenwart: Selbstdarst* (Frankfurt [Main]: Musterschmidt, 1977); *Beyond Existentialism* (Westport, CT: Greenwood Press, 1978).

For further bibliography and evaluation, see Richard Wisser, "Wertwirklichkeit und Sinnverständnis. Gedanken zur Philosophie von Fritz-Joachim von Rintelen," in *Sinn und Sein: Ein philosophisches Symposium F. J. v. Rintelen gewidmet*, edited by Richard Wisser (Tübingen: Niemeyer, 1960), pp. 611–708.

**L. E. Loemker (1967)**

*Bibliography updated by Michael J. Farmer (2005)*

## RITSCHL, ALBRECHT BENJAMIN

(1822–1889)

Albrecht Benjamin Ritschl, the German theologian, was born in Berlin and studied theology at Bonn, Halle, Heidelberg, and Tübingen. He taught theology at Bonn from 1846 to 1864, and at Göttingen for the remainder of his career. Ritschl reexamined Christianity in the light of neo-Kantianism and historicist principles. After 1875 his influence was widespread in a number of German universities and led to increased interest in religious psychology, comparative religion, and related fields. However, his school came under sharp criticism from orthodox, pietist, and liberal quarters.

Ritschl undertook to establish Christian theology as an autonomous and systematic discipline. To do this he had first to purge German religious thought of pietism, Hegelian speculative theism, and the pantheism of Friedrich Schleiermacher and then to apply the techniques and results of contemporary literary and historical criticism. On the basis of Immanuel Kant's ascription of priority to practical reason over theoretical reason and his separation of philosophy and religion, Ritschl distinguished between value judgments and theoretical judgments. Unlike Kant, however, Ritschl accorded primacy to religion over philosophy on the grounds that spirit (the noumenal) takes precedence over matter (the phenomenal); also unlike Kant, he accorded moral primacy to the community (the nation) over the individual.

Ritschl believed that the deep-rootedness and continuity of religion, as expressed in dogmas and institutions, testifies to the reality and superiority of the religious need of practical reason in human nature. This need arises out of a basic contradiction between nature and spirit in human nature. The value of religion and particularly of Christianity, Ritschl thought, can be verified by history, which shows that this contradiction seeks a resolution in some form of redemption in the world. The Kantian elements in Ritschl's thinking, in combination with this positivist tendency, led him to believe that history does not merely provide material in support of some arbitrary, nonhistorical preconception but reveals an essential structure of human consciousness and the intrinsic historicity of Christianity.

In attempting to satisfy both the requirements of history and the claims of practical reason, Ritschl adopted the dogmas of redemption and the kingdom of God as embodied in the life of Christ as the pivots of his religious theory. Man seeks to realize his destiny here on earth by

leading an ethically self-conscious life, which is the core of religiosity. The acts of love that he performs, the content of the ethical life, represent the human counterpart to redemption, and the community required for their performance represents the terrestrial counterpart to the kingdom of God. God's purpose is thus manifest in history. Sin, which is only the result of ignorance, is pardonable because it is only a transitory opposition to this purpose. Ritschl therefore rejected the dogma of original sin as unhistorical and hence unverifiable.

Biblical exegesis led Ritschl to believe that the community is both logically and chronologically prior to the church. Only in the community can man find justification and reconciliation in God. Christ was founder of a community and can be comprehended historically only through our knowledge of how that community conceived him.

From his conviction that religious consciousness is universal and characterized by its quest for redemption, Ritschl concluded that Christianity is the superior expression of that consciousness. History, rather than dogma, verifies Christianity, but its validity is thereby strengthened, not relativized.

Although the community takes precedence over the individual, the individual is not thereby depreciated but is provided with a field within which he is able to realize his personality. While the community is prior to the church, this does not devalue the church's interests but emphasizes its actual efficacy as the ecclesiastical form of the community's organization. Religious truths are established in practice rather than by their appearance in the New Testament, but its authority is thereby strengthened, not subverted. Martin Luther is the most significant religious figure since Christ, not because he modernized Christianity but because he recaptured and restored an understanding of the original Christian attitude.

**See also** Historicism; Kant, Immanuel; Neo-Kantianism; Schleiermacher, Friedrich Daniel Ernst.

### Bibliography

#### WORKS BY RITSCHL

- Die christliche Lehre von der Rechtfertigung und Versöhnung*, 3 vols. Bonn, 1870–1874. Vols. I and III translated (Vol. I by J. S. Black, Vol. III by H. R. Mackintosh and A. B. Macauley) as *Critical History of the Christian Doctrine of Justification and Reconciliation*. Edinburgh and New York, 1872–1900.
- Geschichte des Pietismus*, 3 vols. Bonn: Marcus, 1880–1886.
- Theologie und Metaphysik*. Bonn, 1881.

#### WORKS ON RITSCHL

- Barth, Karl. *Die protestantische Theologie im 19. Jahrhundert*. Zürich: Evangelischer Verlag, 1947. Translated by B. Cozens as *Protestant Thought: From Rousseau to Ritschl*. New York: Harper, 1959. Pp. 390–399.
- Ecke, G. *Die theologische Schule Albrecht Ritschls*, Berlin: Reuther and Reichard, 1897–1904.
- Flügel, O. *Ritschls philosophische Ansichten*. Langensalza, 1886; 3rd ed., 1895.
- Garvie, Alfred E. "Ritschlianism." In *Encyclopedia of Religion and Ethics*, edited by James Hastings, Vol. X, pp. 812–820. Edinburgh and New York, 1919.
- Häring, Theodor. *Zu Ritschls Versöhnungslehre*. Zürich, 1888.
- Mackintosh, R. *Albrecht Ritschl and His School*. London: Chapman and Hall, 1915.
- Ritschl, Otto, *Albrecht Ritschls Leben*, 2 vols. Freiburg and Leipzig: Mohr, 1892–1896.
- Schoen, H. *Les origines historiques de la théologie de Ritschl*. Paris, 1893.
- Weber, W. *Die Frage der Rechtfertigung in der Theologie Albrecht Ritschls*. Heidelberg, 1940. Dissertation.
- Wendland, J. *Albrecht Ritschl und seine Schüler*. Berlin, 1899.

**Robert Anchor (1967)**

## ROBINET, JEAN-BAPTISTE-RENÉ (1735–1820)

Jean-Baptiste-René Robinet, the French *littérateur* and speculative philosopher, was born in Rennes. He started to become a Jesuit, but withdrew from the order and went to Holland to devote himself to letters. There he published his principal work, *De la nature* (4 vols., Amsterdam, 1761–1768), and in 1768, *Considérations philosophiques de la gradation naturelle des formes de l'être, ou les Essais de la nature qui apprend à faire l'homme* (2 vols., Amsterdam and Paris). He eked out an existence by hackwork, translating English novels and giving English lessons. He became embroiled with Voltaire by selling the manuscript of *Lettres secrètes* for publication without Voltaire's permission. He went to Paris in 1778 when he was made royal censor and secretary to one of the king's ministers. During the Revolution he returned to Rennes, where he lived quietly. In addition to many minor pieces, he published a translation of David Hume (*Essais de morale, ou Recherches sur les principes de la morale*, 1760) and edited a vast compilation, *Dictionnaire universel des sciences morale, économique, politique et diplomatique* (London, 1777–1783, 30 vols. in quarto).

*De la nature* caused some stir because of its strange ideas. When it was attributed to François-Vincent Toussaint, Denis Diderot, and Claude-Adrien Helvétius, Robi-

net admitted his authorship in a letter to the *Journal des savants*. The many quotations in *De la nature* testify to its author's vast readings; his thinking, however, is original. It is characterized by a curious mélange of mysticism and scientific spirit. *De la nature* touches on many subjects, but its announced theme is a modern version of Manichaeism: There is an equilibrium of good and evil in all substances and their modes. Robinet's purpose is to exculpate God and establish the necessity of evil. Embracing Benedict de Spinoza's principle that all possibles exist, he attacks Gottfried Wilhelm Leibniz by asserting that, therefore, there can be only one world and that God had no choice in the matter. "God no more had the power to modify the nature of the world than his own nature."

Robinet argued that behind the apparently random distribution of pleasure and suffering in the world there lies a fluid but fixed order. "The physical economy is such that good and evil are engendered with equal fecundity. They flow naturally from the depth of essences." God can in no way remove evil, for omnipotence does not extend to impossibles or contradictions. The suppression of evil implies contradiction, for good without evil would be infinite. The total quantity of good and evil is at every moment equal. Thus the harmony of the world is always the same, and progress is a myth or an illusion. Despite this equilibrium, God is good and his justice is seen in his not having favored one species at the expense of the others; for man is not king of the universe, as Buffon had claimed, and nothing has been created especially for his use. For human beings, life is a balance of happiness and unhappiness, and they should therefore console themselves by the enjoyment of pleasures. Moderation is the best path in all areas of life. The lower classes must be kept in ignorance, for their own benefit and that of the state; slavery is justifiable. Human nature being what it is, equality and fraternity are impossible.

The universe, for Robinet, is animate. All forms of being, including planets and stars, have the power of reproduction. The individual is unimportant, an instrument nature uses for its procreative purposes; only the species endures. Robinet speculates that nature has developed variations on a single prototype; from stones to men, there is a natural gradation of beings. The "prototype" is "a germ which tends naturally to develop itself.... Its energy cannot be repressed.... The germ develops, then, and each degree of development gives a variation of the prototype, a new combination of the original plan." The only difference between stone, plant, and animal is "the measure in which they participate in that essence.... A stone, an oak, a horse are not men; but they can be

regarded as more or less rough types in their relation to a single primitive design." We must consider the succession of individuals "as so many steps of being [advancing] toward humanity."

Robinet draws close to an evolutionary hypothesis in his concept of nature as experimenting and as developing toward greater complexity; he also considers all species as related. It is not a true evolutionism, however, inasmuch as each trial in the ascending scale of complexity is made de novo from the relatively unorganized stage of the original prototype. Species do not themselves have a history but are fixed once they are spewed forth. Robinet also pictures a biological struggle for existence and a natural balance, but does not relate these to transformism. Robinet's work influenced both Johann Gottfried Herder and G. W. F. Hegel and was considered of interest in the former Soviet Union.

*See also* Evolutionary Theory; Diderot, Denis; Evil; Hegel, Georg Wilhelm Friedrich; Helvétius, Claude-Adrien; Herder, Johann Gottfried; Hume, David; Leibniz, Gottfried Wilhelm; Mani and Manichaeism; Spinoza, Benedict (Baruch) de; Voltaire, François-Marie Arouet de.

### **Bibliography**

Damiron, M. *Mémoires pour servir à l'histoire de la philosophie au XVIII<sup>e</sup> siècle*, Vol. II. Paris: Ladrance, 1858.

Rosso, C. *Moralisti del bonheur*. Turin, 1954.

L. G. Crocker (1967)

## ROCHEFOUCAULD, DUC FRANÇOIS DE LA

*See La Rochefoucauld, Duc François de*

## ROHAULT, JACQUES (1620–1672)

Jacques Rohault was a mechanistic Cartesian experimental physicist. He was born in Amiens, France, and earned his MA in Paris in 1641. There, he became Claude Clerselier's Cartesian disciple and son-in-law. He was Pierre-Sylvain Régis's teacher and converted him to Cartesianism. In the 1650s Rohault was a private tutor in Paris, and his "Cartesian Wednesday" evening lectures, complete with laboratory table demonstrations, were

attended by many members of the noble class, women as well as men, and did a great deal toward popularizing Cartesianism. His *Traite de physique* (Paris, 1671) was a standard text for nearly fifty years. John Clarke and Samuel Clarke, rather than writing a Newtonian physics, translated Rohault's work into Latin (1697) and English (1723) and added Newtonian footnotes to correct Rohault's Cartesian mistakes. The *Traite* contains descriptions of explanations and experiments in support of Cartesian mechanistic physics. Like René Descartes, Rohault holds that these explanations are only probable because absolute certainty is unattainable by humans.

Also in Paris in 1671 Rohault published his *Entretiens sur la philosophie*, in which he defends the thesis that Cartesian principles and Christian doctrines do not conflict because each pertains to a separate and distinct realm of truth and knowledge. The book was popular, but Rohault's position was generally viewed as heretical by the Catholic Church.

Rohault opposes Nicolas Malebranche's occasionalism and presents his own mechanistic Cartesianism based on eight axioms he takes to be self-evident:

- (1) Nothing (that which has no existence) has no properties
- (2) Something cannot possibly be made of nothing, that is, nothing cannot become something
- (3) No thing or substance can be annihilated, that is, something cannot be reduced to nothing
- (4) Every effect presupposes some cause
- (5) If one does not cause an effect, that effect necessarily depends on some other cause
- (6) Everything endeavors to continue in the state in which it is (an early Cartesian rendering of a principle of inertia)
- (7) Every alteration is made by some external cause, that is, in opposition to Aristotle, no material thing can alter itself through an inner power, force, or form
- (8) Every alteration is proportional to the force of the causal agent

Certain propositions follow logically from these axioms, but Rohault says these truths of reason remain purely formal and have no application if there are no existents. Thus, the first task in understanding the world is to seek out existents. In strict Cartesian order one knows first one's own self, whose existence Rohault proves syllogistically:

- (a) From principle (1) above, whatever has properties is something
- (b) Thinking is a property
- (c) Whatever thinks, therefore, exists as something because it has the property of thinking
- (d) I think
- (e) Therefore, I exist

Reasoning with these principles about ideas and sensations leads to knowledge of the essences of mind, God, and matter and to proofs of the existence of God and of matter. The essence of mind is thought; of God, necessary existence; and of matter, extension. Rohault states that mind and matter are completely different but that God so created the human mind or soul such that motions caused by material impressions on the sense organs and in the brain of the body with which it is united give rise in the soul to sensations and ideas. Neither sensations nor ideas resemble material things, and so resemblance is not necessary for knowledge. It is simply the nature of sensations to give knowledge of the existence of material things, and the nature of some ideas is to give knowledge of the place, situation, distance, magnitude, figure, number, and motion or rest of material things.

Rohault's method in physics is to reason mathematically about experiments before conducting them. His goal is to explain the sensible effects of material things. For this only the primary material properties of size, figure, motion, and arrangement of divisible, impenetrable particles in a plenum are needed; occult qualities such as Aristotelian forms are unnecessary.

In *Entretiens de philosophie* (Paris, 1671), the companion volume to the *Traite*, Rohault explains in mechanical terms Cartesian opinions on animal machines and transubstantiation. Animal behavior, he claims, can be explained if animals are completely material; human behavior, however, requires a rational soul that is immaterial, hence indivisible, hence immortal. For Cartesians, the sensible qualities as they exist in material things are not seen, tasted, and so on as one sees and tastes them, but are merely the powers bodies have, determined by the size, figure, motion, and arrangement of their particles, to cause sensations in the mind. There is no further explanation of these powers beyond the fact that God made the correlations between bodily movements and one's sensory experience.

Transubstantiation, then, is the point-by-point replacement of bread and wine by Christ's flesh and blood. Therefore, the flesh and blood of Christ that occu-

pies the places (is bound by the surfaces) formerly occupied by bread and wine causes sensations exactly like those that the bread and wine formerly caused. Consequently, real accidents or Aristotelian forms subsisting separately from Aristotelian matter as postulated in the scholastic explanation of transubstantiation are unnecessary. There are further physical explanations and assurances that Cartesian principles do not contradict Catholic doctrine in *Oeuvres posthumes de Rohault* (Paris, 1682).

Overall, Rohault disclaims metaphysics and says that although the substitutions are miraculous, even his mechanist explanation of transubstantiation is only a solution to a problem in physics. His work illustrates the strong empiricist stress on observation and experiment toward probable mechanistic explanations in physics so prominent in many Cartesian philosophers. Finally, use of Rohault's *Traite* as a physics textbook merely with addition of Newtonian footnotes constitutes a major shift to nonmetaphysical, explanatory concerns in science.

**See also** Aristotelianism; Aristotle; Cartesianism; Clarke, Samuel; Descartes, René; French Philosophy; Malebranche, Nicholas; Régis, Pierre-Sylvain.

### **Bibliography**

- Schmaltz, Tad M. *Radical Cartesianism: The French Reception of Descartes*. New York: Cambridge University Press, 2002.
- Watson, Richard A. *The Breakdown of Cartesian Metaphysics*. Indianapolis, IN: Hackett, 1998. Originally published as *The Downfall of Cartesianism, 1673–1712: A Study of Epistemological Issues in Late Seventeenth-Century Cartesianism* (The Hague: Nijhoff, 1965).

*Richard A. Watson (1967, 2005)*

## ROMAGNOSI, GIAN DOMENICO

(1761–1835)

Gian Domenico Romagnosi was born in Salsomaggiore, near Parma, and studied at the Collegio Alberoni in Piacenza. Through the teaching of Giovanni Antonio Comi, a follower of Gottfried Wilhelm Leibniz and Christian Wolff, Romagnosi became acquainted with the doctrines of Étienne Bonnot de Condillac and with the writings of Charles Bonnet, which had a decisive influence on him. After his graduation in 1786, he conceived his best-known work, *Genesi del diritto penale* (Genesis of Penal Law; completed in 1789 and published in Pavia in 1791), in which he claimed that the fundamental right to punish

belongs to society. Society alone, and not the individual, can mete out “that amount of evil that is necessary to preserve the well-being of our fellow men” and can oppose the “criminal impulse” with a “moral counterimpulse.”

Named mayor of Trent in 1791, Romagnosi remained in that office for ten years, during the period of the French Revolution and the rise of Napoleon Bonaparte. During this time he published, among other works, his *Cosa è l'eguaglianza* (What Is Equality?; Trent, 1792) and *Cosa è libertà* (What Is Freedom?; Trent, 1793). After a brilliant political career under the Napoleonic government, he became professor of natural and public law at Parma (1802), but after the restoration he was dismissed from his position and was arrested. The Austrian government also prevented him from accepting a post at the Ionian University at Corfu offered to him by Frederick North, Lord Guilford. Regarded as a master by Italian patriots, Romagnosi died, after a sad but active old age, in Milan. His major works, in addition to the *Genesi*, are considered to be the *Introduzione allo studio del diritto pubblico universale* (Introduction to the Study of Universal Public Law; Parma, 1805), the *Assunto primo della scienza del diritto naturale* (A First Thesis on the Science of Natural Law; Milan, 1820), and a series of essays on *incivilimento* (civilizing, or the process of civilization) in 1832.

Although he was influenced by the Enlightenment, Romagnosi remained attached to the historicism of Giambattista Vico and followed a “positive” method of research, advocating the activity of the human spirit rather than sensationalism and substituting for the abstractness of the isolated human individual the concreteness of the nation as the subject of the historical action. In epistemology he refused to reduce all cognitions to “transformed sensations,” but at the same time he denied that intelligence is independent of sensitivity: In reality, “discernment” is already present in “feeling.” The mind acts by means of its own “rational signs.” These cannot be regarded as preexisting ideas but, rather, as manifestations of mental activity, which, along with the sensory datum, gives form to experience. By contrast, the correspondence of our prior judgments with the actual signs of things, that is, with experience, constitutes the criterion of the truth of our knowledge, which is sought and found pragmatically.

Romagnosi's civil philosophy is the most interesting part of his work: Man is real only in a historically determined society—the “collective person of the society”—which is in a state of constant civilizing progress and whose characteristic traits, elements, and laws Romagnosi



sought to define. Romagnosi's doctrine of *incivilimento* constituted a philosophy of history faithful to the concrete development of real events, in contrast with that of G. W. F. Hegel, which Romagnosi opposed as "ultrameta-physical." Society develops through the synthesis of national character (tradition) with stimulation—spontaneous, free, and renewing—according to a law of convenience, with all parts of the nation tending toward an equilibrium of force and utility through the balance of interests and powers. This dialectic of civilization is a work of art, even the highest work of art of a humanity striving for perfection.

**See also** Bonnet, Charles; Condillac, Étienne Bonnot de; Enlightenment; Hegel, Georg Wilhelm Friedrich; Historicism; Leibniz, Gottfried Wilhelm; Punishment; Vico, Giambattista; Wolff, Christian.

### Bibliography

Editions of Romagnosi's works are *Opere*, 19 vols. in 24 (Florence, 1832–1837), and *Opere*, 8 vols. in 16 (Milan: Perelli e Mariani, 1841–1848; actually 1841–1852), both edited by Alessandro de Giorgi; *Opere scelte*, 3 vols. (Rome: Reale accademia d'Italia, 1936–1937); and *Opere di G. D. Romagnosi*, C. Cattaneo, G. Ferrari, edited by Enrico Sestan (Milan: Ricciardi, 1957).

Works on Romagnosi are L. G. Confalerione, *G. D. Romagnosi. Notizie storiche e biografiche. Bibliografia e documenti* (Carate Brianza, 1928); Achille Norsa, *Il pensiero filosofico di G. D. Romagnosi* (Milan, 1930); G. A. Belloni, *Romagnosi, profilo storico* (Milan, 1935); E. di Carlo, *Bibliografia Romagnosiana* (Palermo, 1936); Francesco Orestano, *Giandomenico Romagnosi* (Rome, 1935); Giacomo Perticone, *La filosofia "civile" di G. D. Romagnosi* (Florence, 1935); Alessandro Levi, *Romagnosi* (Rome, 1935); Lorenzo Caboara, *La filosofia politica di Romagnosi* (Rome: Istituto di filosofia del diritto, 1936); and Loris Ricci Garotti, "G. D. Romagnosi nella critica recente," in *Società* 14 (1959): 109–140.

**Eugenio Garin (1967)**

*Translated by Robert M. Connolly*

## ROMAN PHILOSOPHY

See *Cicero, Marcus Tullius; Lucretius; Marcus Aurelius Antoninus; Seneca, Lucius Annaeus; Stoicism*

## ROMANTICISM

"Romanticism" and "romantic" are protean words, the despair of a rigorous semanticist. They designate a gener-

ally accepted period, especially in literature and the arts, of Western cultural history, roughly from the late eighteenth to the mid-nineteenth century. They embrace a cluster or syndrome of ideas about the true, the good, the beautiful, philosophical ideas both in the popular and in the technical sense, ideas endlessly debated in the last few centuries. Although the behavioral scientists groping to establish a rigorous classification of human personality generally eschew the word, *romantic* remains in common use to describe a temperament or personality often, perhaps usually, held to be a constitutional element of an individual and at least in part independent of cultural fashion. In all these senses "romanticism" and "romantic" cover a multitude of particulars that in a given combination can appear very different, if not mutually incompatible. Hence so good a historian of ideas as Arthur Lovejoy urges the use of the plural, *romanticisms*, and can write of the "Chinese origins of a romanticism"; and W. T. Jones insists that romanticism can only be understood as a very complex syndrome of "biases" in the direction of what he calls the dynamic, the disordered, the continuous, the soft-focused, the inner, the this-worldly.

### THE ROMANTIC TEMPERAMENT

Sensitive, emotional, preferring color to form, the exotic to the familiar, eager for novelty, for adventure, above all for the vicarious adventure of fantasy, reveling in disorder and uncertainty, insistent on the uniqueness of the individual to the point of making a virtue of eccentricity, the typical Romantic will hold that he cannot be typical, for the very concept of "typical" suggests the work of the pigeonholing intellect he scorns. Though his contempt for this world of reason and commonsense calculation may push him toward otherworldliness, the Romantic is too much a man of words and sensations to make a good mystic. He may admire the mystic, especially the exotic mystic from the East, but he himself is a good Westerner. In fact, the difficulties of reconciling the often contradictory particulars of romanticism in respectable generalization come out in any attempt to isolate a romantic personality. William Blake has most of the marks of the Romantic, from the positive one of extreme transcendental yearning to the almost universal romantic negative one of contempt for the "meddling intellect"; yet in his quite otherworldly drawings his symbolic, mystical figures are delineated with a draftsmanship of classical solidity and of firm this-worldliness. There is nothing fuzzy, nothing Turner-like, in Blake's art. William James has the full romantic love for the struggling, the unestablished, the untried; but he cannot be accused of what he

himself called “tender-mindedness,” of idealistic distrust of the instrument of thought. Friedrich Nietzsche, who used “romantic” as a term of reproach, who said of Richard Wagner’s music that it sweats, and called Mme. de Staël “that prolific ink-yielding cow,” shared all the romantic hatreds for the shopkeeper’s world of grubbing common sense and above all had the Romantic’s desire for *etwas mehr*, the something more of Percy Bysshe Shelley’s “desire of the moth for the star.”

However difficult the romantic personality may be to isolate in analysis, it can be recognized all through Western cultural history, and indeed in the active life of enterprise and politics. Euripides and Catullus were surely Romantics. The *Odi et amo* (I hate and I love) of Catullus is a classic assertion of romantic ambivalence; the *rumoresque senum severiorum/omnes unius aestimemus assis* (Let us regard all the gossip of censorious old men as not worth one penny) is a fine assertion of one of the minor marks of romanticism, contempt for the Philistine decencies of the old in spirit. François Villon and François Rabelais were Romantics, even though they were Frenchmen who, as Frenchmen, so nineteenth-century English and German romanticists thought, should have been incapable of transcending the petty ways of *mesure* and *la raison raisonnée*. In our own day, the romantic temperament crops up everywhere—in artists and poets of course, but also in philosophers. Henri Bergson was a Romantic. But so too, it may be argued, was A. N. Whitehead; and there are scientists not untouched by the desire of the moth for the star. In active life, Alexander the Great and Napoleon Bonaparte were Romantics; Frederick the Great and Otto von Bismarck were classicists.

There are then, in our Western civilization, presumably always born romanticists and born classicists—or born Dionysians and born Apollonians, to use an expressive dualism especially popular with the Germans from Gotthold Ephraim Lessing through Nietzsche to Oswald Spengler. (The Germans usually classify themselves as the great Dionysian force in the West.) We can but guess at the distribution of these two types in a general population. Probably the well-defined or extreme temperaments are limited in numbers always; most human beings can adapt to the fashion of their age. In one age, say Vergilian and Horatian Rome, or the France of Louis XIV, the Apollonian is dominant, the Dionysian subdued, even silent. Sometimes in Apollonian ages, however, the Dionysian is the rebel, the man out of tune with his times; Giambattista Vico, perhaps, should be so listed in the Apollonian early eighteenth century. In another age, and notably in the Romantic Age here considered, the

Dionysian is dominant and the Apollonian repressed, sometimes tempted, as was the quite unecstatic J. S. Mill, to romantic depths of understanding.

## ROMANTICISM AND THE ENLIGHTENMENT

One type can be dominant, but not in sole and exclusive possession. To the cultural historian, the early and mid-eighteenth century and the early nineteenth can stand for two great antithetical styles or fashions: the first, classical or enlightened; the second, romantic. The years from about 1770 to the first decade of the nineteenth century are obviously years of transition. In a graph, the rising lines of Romanticism cross the descending lines of classicism somewhere in the 1770s in Germany (with the heyday of “Sturm und Drang”), 1798 in England (with the publication of the *Lyrical Ballads*), and 1820 in France (with the publication of *Méditations* by Alphonse-Marie-Louis de Prat de Lamartine). But even after the triumph of Romanticism as a cultural fashion, individuals and groups continued to display the tastes and attitudes associated with the classicism and rationalism of the eighteenth-century Enlightenment. J. S. Mill tells us in his autobiography that he was influenced by the lyricism and even the transcendentalism of the Lake poets, notably Samuel Taylor Coleridge; but the influence seems not to have weaned him away from the fundamentals of Benthamite thought. In France the thought of such men as Comte de Saint-Simon, Louis Blanc, Auguste Comte, though some of the externals of romantic fashion are visible among them, is, on the whole, along with that of the French Left generally, true to the traditions of the *philosophes*. Even in Germany, a philosopher such as Ludwig Feuerbach asserts the unromantic doctrines of materialism; and Marxism itself, though it shows romantic marks—the concept of the dialectic, derived of course from G. W. F. Hegel, is essentially romantic in its insistence on change as an overcoming of contradictions—is nonetheless committed to an optimistic and very eighteenth-century stand on the rational organization of man and society.

The romantic generation was indeed very conscious of breaking sharply with its parents and grandparents. Few breaks between cultural generations in the West have been more vigorously asserted than this one. The romantic youth absorbed in the depths of William Wordsworth’s *Prelude*, or Vicomte Chateaubriand’s *Génie du Christianisme*, or Johann Wolfgang von Goethe’s *Faust* felt nothing but contempt for the abstract ideas and the confined tastes of his shallow Voltairian grandfather. To a

surprising extent, the fashionable Romantic was—or claimed to be—in all things the opposite of the Enlightened. Yet our own generation can hardly avoid holding that the romantic rebellion against its parent was in itself a proof of the filial relation between Romanticism and Enlightenment. Not only were the ideas of men like Jean-Jacques Rousseau, Vico, Lessing, and even Denis Diderot, all of whom lived at the height of the Enlightenment, seminal to all later Romanticism, but both Enlightenment and Romanticism shared much—a belief in process, change, if not actually progress, a belief in the possibilities of manipulating the environment, indeed a fundamental and very modern relativism never really transcended in the search for eternal verities. Both, whatever their metaphysical position on the problem of determinism, in practice displayed a firm conviction that things not only change, but that they can be changed by human effort. Of many specific doctrines—primitivism, for instance, or individualism in ethics and politics—it is hard to decide whether they are more characteristic of enlightened or of romantic thought.

#### SOME SPECIFIC ROMANTICISMS: ART AND LETTERS

The romantic touch is extremely visible in all the arts, from painting through architecture to interior decoration. Bright colors, or soft and fuzzy ones; exotic themes, Oriental scenes; crowded and action-filled historical paintings—concretely, almost any canvas by Eugène Delacroix—set romantic painting off from the sculptured Roman figures of David. And yet, to point up the coexistence of the romantic and the classical throughout the period, the sharp outlines, the measured realism—the Romantic would hold, the conventionality—of the portraits by Jean-Auguste-Dominique Ingres, who survived until 1867, outdo David's in classical firmness. The great romantic style in architecture was the neo-Gothic, itself a manifestation of the romantic rehabilitation of everything medieval that had been held in contempt by the Renaissance and Enlightenment. Yet Neo-Gothic was never a dominant style, not even in the Nordic lands; moreover, it soon fell into a most unmedieval and unromantic regularity and repetitiveness of detail. But Romanticism did rescue from the neglect in which they had long been left the great medieval cathedrals. In the decorative arts romantic tastes were extremely eclectic, fond of the exotic, addicted to rich dark woods and, in the climax of the Victorian drawing room, to a clutter of display wholly dependent on the existence of inexpensive domestic labor. In music, the romantic at its extreme

went in for program music, birdcalls and thunderstorms, vast orchestras, and appropriate dissonances. The difference between the music of Joseph Haydn or Wolfgang Amadeus Mozart and that of Hector Berlioz or Wagner, like that between the painting of David and that of Delacroix, is obvious to the most untutored.

Poetry, the novel, and history are the great romantic literary genres, and in all of them the romantic syndrome is readily recognized. Although Goethe was a complex personality who was frequently in conflict with contemporary representatives of the romantic movement, his *Faust* is in itself a masterly summary of romantic themes: revolt against the dullness, the narrowness of rationalism (“gray dear friend is all theory, green only life's eternal tree”), striving for *etwas mehr*, for the infinite (the essential theme of Faust's bargain with Mephistopheles); contempt for the Philistine, the literal-minded ordinary man (the walk with Wagner); primitivism (Gretchen's innocence); ambivalence (“Two souls, alas, live in my breast”); and much else, right on to the final *chorus mysticus* of Part II. Indeed, this last is a fine touchstone; anyone who finds it nonsense or at least unpalatable is definitely not Romantic:

Alles Vergängliche  
Ist nur ein Gleichnis;  
Das Unzulängliche  
Hier wird's Ereignis;  
Das Unbeschreibliche  
Hier ist's getan;  
Das Ewig-Weibliche  
Zieht uns hinan.

The three English Lake Poets, Wordsworth, Coleridge, and Robert Southey, together pretty well cover the romantic range; and Wordsworth's “The Tables Turned” (“One impulse from a vernal wood,” “We murder to dissect,” “Enough of science and of art;/Close up those barren leaves”) states the central position of the romantic *Weltanschauung* almost as neatly as Goethe's *Gefühl ist alles*. One more figure, one more complex of themes, is needed to round out our concept of romantic poetry: This is the unhappy, misunderstood, heroic Promethean figure, half Shelley and all Lord Byron. In terms of sheer educated fashion, Byron and his whole train of European congeners (imitators would be an unfair word here)—Alfred de Musset, Alfred-Victor de Vigny, Giacomo Leopardi, José de Espronceda, Mikhail Lermontov, and the rest—may stand for the romantic poet.

Forerunners of the romantic novel are clear in the eighteenth-century “Gothic” novel, such as those of Ann Radcliffe (so charmingly satirized by the nonromantic

Jane Austen in *Northanger Abbey*); in the sentimental novel, such as Rousseau's *Nouvelle Héloïse*; and in the psychological novel of disturbed and disturbing love, such as Pierre-Ambroise-François Choderlos de Laclos's *Liaisons dangereuses* and the novels of the Marquis de Sade. The psychological novel reaches its best in the work of Stendhal, whose heroes foreshadow a long line of adventurers of soul and body, a line by no means extinct today. Yet in terms of the wider public of romantic fashion, Walter Scott's Waverley novels were the great success of their day. They carried their audience back into a simpler, more varied, more interesting past than the present of the Industrial Revolution. They exemplified that other inheritance from the German side of the Enlightenment, the theme, best marked in Johann Gottfried Herder, of organic historic growth of a folk spirit, a folk character, a product of time, not a product of the planning, present-bound intellect. One lost one's self in Sir Walter's pages, became one with one's own best past. We are a long way from Henry St. John Bolingbroke's definition of history as "philosophy teaching by examples."

#### HISTORY AND POLITICAL THOUGHT

The writing of serious history received a great impetus from the romantic movement, and in particular from Scott's work. Augustin Thierry, Jules Michelet, the Heidelberg school in Germany; in England Henry Hallam, indeed T. Macaulay, by no means a Romantic in temperament; and in the United States the great New England school of W. H. Prescott, J. L. Motley, and Francis Parkman wrote history for a wide reading public, history with narrative force and movement, history with a message of patriotism, of identification with a folk, yet also history carefully reconstructed by painstaking research. The historian and the critic of art and literature insisted on one of the great romantic themes: continuity, the continuity of life and flow, growth, development; a process, to the Romantic, always denatured, indeed destroyed, by the dividing analytical mind ("We murder to dissect").

The complexities and difficulties of generalizing about Romanticism come out most clearly, perhaps, in the field of political thought. You can, of course, always construct a pair of Procrustean beds: a conservative bed for the Romantics; a liberal, progressive, or radical bed for the Enlightened. Edmund Burke and Scott can be squeezed into the first, Thomas Paine, W. Godwin, Thomas Jefferson into the second. But the trouble is that you can quite plausibly switch the beds, putting the Romantics into the liberal or progressive bed, the Enlightened into the conservative bed. Shelley, Byron,

Benjamin Constant can go into the first; Voltaire (surely no democrat), John Adams, the *idéologues* who rallied to Napoleon can go into the second. But Victor Hugo would have to be divided, his younger self put into the conservative, his older self into the liberal bed.

Critics have indeed tried to fix Romanticism on one side or the other in politics, and—given their premises—not without some success. Probably in the balance Romanticism has worked toward the growth of modern democracy, toward a belief in progress and toward "liberty, equality, fraternity," toward the open society—toward much, in fact, that gets its start from the rationalists of the Enlightenment. Yet the Burkean belief in human fallibility, human blindness of passion, and in tradition-enshrined institutional dikes to restrain these anarchic thrusts (dikes not to be tampered with by the intellect), as well as belief in the folk, in an organic society not the product of planning, is surely also congruous with much of Romanticism. So too is the anti-intellectual strain that comes out much later in theories of racism, elitism, *Blut und Boden*, in Nazism and Fascism.

#### PHILOSOPHY

Romanticism is more than a fashion in arts and letters, more than an approach to political problems: It is a philosophy, or better, a set of philosophies loosely tied together if only by their common rejection of eighteenth-century rationalism, of refusal to line up, shall we say, on the Locke-Hume axis. Arthur Schopenhauer is the arch Romantic, the extreme Romantic, among formal philosophers. The world of phenomena, of sense perception, is to him unreal; the will that moves the universe is real enough, but certainly is not rationally knowable by those it moves; this will is blind, shapeless, evil; life, merely phenomenal though it is, is still for us all painful, wearisome, a long unhappy voyage (note the metaphors of movement); Schopenhauer seems at times to hold that a nirvana of surcease is perhaps attainable; at any rate, this life is hopeless.

Romantic pessimism is not, however, the central theme of philosophy in these years. Hegel, at bottom an optimist, is much more central. In a sense, the great romantic philosophers, most of them Germans, go back to Immanuel Kant, who always thought of himself as firmly enlightened, and whose brief *Was ist Aufklärung?* is one of the landmarks of the century of prose and reason. The romantic seedling in Kant, however, is his distinction between the noumenal and the phenomenal, and his resolution of the dualism by what amounts to intuition or faith. Johann Gottlieb Fichte and Friedrich Schleierma-

cher and the rest developed this essentially romantic reliance on a “faculty” transcending common calculating logic. Hegel accepted, and gave his own turn to, this very old dualism of spirit-matter, real-unreal, and sought to bring them together by his famous and influential concept of the dialectic of thesis–antithesis–synthesis. The dialectic in all its forms displays a most nineteenth-century and romantic general bias toward historicism, process, development—but such a process seen teleologically as an end, a purpose. For Schopenhauer, there was no end save extinction. But for Hegel there was an end, a vague one, a Germanic eternal peace in which change somehow turns out to be, in the workings of the World-Spirit, the real form of permanence.

These philosophers, trained and subtle professionals whom we have no doubt traduced in this brief account, are less definitely to be associated with Romanticism as a broad cultural movement than the popularizers, the essayists, the preachers. To many devotees of Thomas Carlyle, Ralph Waldo Emerson, and John Ruskin, and to those who listened to bumblerers like Bronson Alcott, romantic philosophy became fashionable transcendentalism, an agreeable summary of the less difficult phases of romantic thought—contempt for the rationalist side of the eighteenth century (indeed, blindness to the existence of any other side of that century), exaltation of intuition, spirit, sensibility, imagination, faith, the unmeasurable, the infinite, the wordless—or at least, only the noblest sounding words. This sort of Romanticism was indeed a solace and an escape, an escape from the difficult and unlovely works that science, technology, and industry were building. But it is by no means the whole of Romanticism, which as a spiritual spur to precisely the kind of invention, adventure, and enterprise, to the preoccupation with change and growth, that was building the new world of the nineteenth century, must be seen as having played, and as continuing to play, an essential part, along with the rational and scientific inheritance from the eighteenth-century Enlightenment, in building our own world of today.

**See also** Art, Expression in; Bergson, Henri; Blake, William; Bolingbroke, Henry St. John; Burke, Edmund; Carlyle, Thomas; Chateaubriand, François René de; Coleridge, Samuel Taylor; Comte, Auguste; Diderot, Denis; Emerson, Ralph Waldo; Enlightenment; Feuerbach, Ludwig Andreas; Fichte, Johann Gottlieb; Godwin, William; Goethe, Johann Wolfgang von; Hegel, Georg Wilhelm Friedrich; Herder, Johann Gottfried; Hume, David; James, William; Jefferson, Thomas; Leopardi, Count Giacomo; Lessing, Gotthold Ephraim;

Locke, John; Lovejoy, Arthur Oncken; Mill, John Stuart; Neo-Kantianism; New England Transcendentalism; Nietzsche, Friedrich; Paine, Thomas; Pessimism and Optimism; Rousseau, Jean-Jacques; Ruskin, John; Saint-Simon, Claude-Henri de Rouvroy, Comte de; Schleiermacher, Friedrich Daniel Ernst; Schopenhauer, Arthur; Shelley, Percy Bysshe; Spengler, Oswald; Staël-Holstein, Anne Louise Germaine Necker, Baronne de; Vico, Giambattista; Whitehead, Alfred North.

### **Bibliography**

Romanticism is not only a complex cluster of ideas; it is one that arouses strong feelings among critics and historians, and that has had its ups and downs in the estimation of the various cultural generations since the late eighteenth century. The following should set the reader on his way through these thickets of critical and philosophical discussion of Romanticism.

Howard Hugo, ed., *The Romantic Reader* (New York: Viking Press, 1957), is an admirable anthology with a good bibliography of works in English and a useful prologue, “What the Romantics Said about Romanticism.” W. T. Jones, *The Romantic Syndrome* (The Hague: Nijhoff, 1961), presents a very suggestive analysis, helpful for all study of the history of ideas. Jacques Barzun, *Classic, Romantic and Modern* (Boston: Little, Brown, 1961), contains the ablest defense of Romanticism; see the section “Romantic—A Sampling of Modern Usage” (pp. 155–168). G. A. Borgese, “Romanticism,” *Encyclopaedia of the Social Sciences* (New York, 1934), Vol. XIII (VII), a remarkably rich brief account, with full bibliographies up to 1934 in all Western tongues, is sympathetic. Irving Babbitt, *Rousseau and Romanticism* (Boston: Houghton Mifflin, 1919), is still the sharpest attack on Romanticism. A. O. Lovejoy, *Essays in the History of Ideas* (Baltimore: Johns Hopkins University Press, 1948), contains several pertinent essays, especially one titled “On the Discrimination of Romanticisms.” Sir Maurice Bowra, *The Romantic Imagination* (Cambridge, MA: Harvard University Press, 1949), is a graceful essay by a distinguished English scholar and critic. Walter Jackson Bate, *From Classic to Romantic* (Cambridge, MA: Harvard University Press, 1946), is one of the best studies of the complex interweaving of classic and romantic in English literature. Ricarda Huch, *Blütezeit der Romantik*, 12th ed. (Leipzig, 1922), and *Ausbreitung und Verfall der Romantik*, 10th ed. (Leipzig, 1922), are sympathetic and graceful accounts of the German Romantics. Pierre Lasserre, *Le Romantisme français* (Paris, 1907), is an unsympathetic account of the French Romantics.

*Crane Brinton (1967)*

## ROMANTICISM [ADDENDUM]

When romanticism is understood broadly, as referring to a major development in European thought and culture

since the turn of the nineteenth century that shows itself distinctly in the spheres of art, historical writing, and political thought, the concept has only a limited role to play in the history of philosophy: Certain very general notions—an emphasis on agency, expression, the cognitive dimension of affect, and the potential of human beings to become genuine wholes—can be described as manifestations of romanticism in philosophy, but the term does not serve to pick out any more determinate set of philosophical commitments.

Here, as with modernism, is a category that is indispensable for general intellectual history, but lacks equivalent value in the history of philosophy. Where the concept does achieve significant purpose in the history of philosophy is in its much narrower application to the group of thinkers based in Jena at the very end of the eighteenth century known as the (early) German romantics, or *Frühromantik*, whose activity centered on production of the *Athenäum*, a journal whose historical importance far exceeds its short life span. Friedrich von Schlegel and Novalis (the pen name of Friedrich von Hardenberg) comprise the philosophical core of German romanticism, with F. W. J. von Schelling and F. D. E. Schleiermacher in close, albeit temporary and qualified, association. J. C. F. Hölderlin—like Novalis, a major German lyric poet—did not belong to the group in Jena but is considered properly as belonging to the same philosophical tendency as Schlegel and Novalis.

Philosophical understanding of the German romantics has been obstructed by the fragmentary form of much of their output, and the literary concern of the movement taken as a ground for assuming its importance to lie outside philosophy, but more recent work, above all by Manfred Frank (b. 1945), Ernst Behler (b. 1928), and Frederick Beiser (b. 1949), has revealed the distinctiveness and importance of the philosophical outlook formulated by the German romantics in the context of the problems and issues facing post-Kantian philosophy. The problems of Immanuel Kant's legacy revolved in the first place around the perceived incompleteness of Kant's transcendental or critical philosophy, which was considered to have opened up a new range of intellectual possibilities and yet to require further development for it to fulfil its emancipatory promise and thereby meet the demands of the age. Johann Gottlieb Fichte's attempt to do exactly this in his *Wissenschaftslehre* held the attention of, but failed to convince, the German romantics, who accepted the rational necessity of seeking to construct a self-grounding philosophical system but believed themselves

to have achieved insight into the reasons why this ideal cannot be realized and must remain an infinite task.

Schlegel's original and influential conception of irony as not merely a literary trope, but rather a corollary of the structure of reflection that, having achieved critical freedom, cannot bring itself to a halt, was developed in part to rationalize this complex attitude toward the ideal of philosophical systematicity. The reorientation proposed by the German romantics in place of Fichte's *Wissenschaftslehre* centered on a novel and very high valorization of art and the aesthetic. This move, far from signalling an aestheticist turning away from the philosophical tasks and the social and political realities that occupied Kant and Fichte, was envisaged as engaging with the full spectrum of philosophical, practical, and cultural problems. The key to the importance ascribed by the German romantics to art—at least, to that art which possesses the qualities of what they called *Poesie* (*romantische*)—lay in its supreme exemplification of true (organic) unity, its synthetic relation to the metaphysical oppositions that structure human existence and reality at large, and its embodiment or symbolization of freedom.

Both this conception of art and the romantics' claim for its practical importance show the influence of Schiller's *Letters on Aesthetic Education*, but the German romantics projected the concept of art and the aesthetic, in a way that Schiller had not, well beyond the sphere of works of art in the strict sense. Schelling's account in the final part of his *System of Transcendental Idealism* of art as what he calls the only true organ and document of philosophy provided one formulation of the German romantic idea that art is philosophically preeminent, and the *Naturphilosophie* that Schelling developed in the late 1790s, which attributes organic status to nature as a whole, and disputes the primacy of mechanism over teleology maintained in modern philosophy even by Kant, falls equally into line with the German romantic program.

In the ethical and political sphere, the German romantics sought to achieve recognition for the claims of personal individuality while at the same time urging the pursuit of organic wholeness in collective life, in opposition to Kant's ethical universalism and political atomism, yet without any intention of contradicting Kant's modern affirmation of freedom. Schleiermacher's ethical theory, though it was composed some years after the dissolution of the romantic circle, may be regarded as giving systematic shape to at least some German romantic ethical insights, in the same way that his earlier, highly successful

work, *On Religion*, stood in close accord with the German romantic intention to recreate religious forms.

German romanticism has affinities with positions that had been developed earlier by J. G. Hamann, J. G. Herder, and F. H. Jacobi in their critique of the German Enlightenment, and some commentators have suggested that it also prefigures deconstruction and postmodern philosophy, on account of its skepticism regarding the attainability of final philosophical truth. This view runs a risk of anachronism, however, for while it is true that German romanticism diverges from the three great developments of German idealism, it nevertheless remains committed to an ideal of rationality and retains many of the idealistic, not to say Platonistic, elements that are present in the philosophies of Fichte, Schelling, and Hegel. Indicative in this regard is the fact that the sharp criticism made by Hegel of German romanticism—which collapses, Hegel believes, into hyper-subjectivism and arbitrariness—is premised on an understanding of the movement as having grasped, without giving adequate form to, important philosophical truths.

**See also** Enlightenment; Fichte, Johann Gottlieb; German Philosophy; Hamann, Johann Georg; Hegel, Georg Wilhelm Friedrich; Herder, Johann Gottfried; Hölderlin, Johann Christian Friedrich; Jacobi, Friedrich Heinrich; Kant, Immanuel; Modernism and Postmodernism; Novalis; Schelling, Friedrich Wilhelm Joseph von; Schlegel, Friedrich von; Schleiermacher, Friedrich Daniel Ernst.

## Bibliography

### COMMENTARY

- Behler, Ernst. *German Romantic Literary Theory*. Cambridge, MA; New York: Cambridge University Press, 1993.
- Beiser, Frederick. "Absolute Idealism." In *German Idealism: The Struggle Against Subjectivism 1781–1801*. Cambridge, MA: Harvard University Press, 2002.
- Beiser, Frederick. *The Romantic Imperative: The Concept of Early German Romanticism*. Cambridge, MA: Harvard University Press, 2003.
- Craig, Edward. "The Metaphysics of the Romantic Era." In *The Mind of God and the Works of Man*. Oxford, U.K.: Clarendon Press, 1987.
- Frank, Manfred. *The Philosophical Foundations of Early German Romanticism*. Translated by Elizabeth Millán-Zaibert. Albany: State University of New York Press, 2004.
- Frank, Manfred. "Unendliche Annäherung": *Die Anfänge der philosophischen Frühromantik*. Frankfurt am Main: Suhrkamp, 1997.
- Henrich, Dieter. *Course of Remembrance and Other Essays on Hölderlin*, edited by Eckart Förster. Stanford, CA: Stanford University Press, 1997.

- Lacoue-Labarthe, Philippe, and Jean-Luc Nancy. *The Literary Absolute: The Theory of Literature in German Romanticism*. Translated by Philip Barnard and Cheryl Lester. Albany: State University of New York Press, 1988.
- Richards, Robert J. *The Romantic Conception of Life: Science and Philosophy in the Age of Goethe*. Chicago: University of Chicago Press, 2002.
- Seyhan, Azade. *Representation and Its Discontents: The Critical Legacy of German Romanticism*. Berkeley: University of California Press, 1992.

### WRITINGS IN TRANSLATION

- Beiser, Frederick, ed. *The Early Political Writings of the German Romantics*. Cambridge, MA; New York: Cambridge University Press, 1996.
- Bernstein, J. M., ed. *Classic and Romantic German Aesthetics*. Cambridge, U.K.; New York: Cambridge University Press, 2003.
- Hölderlin, Friedrich. *Friedrich Hölderlin: Essays and Letters on Theory*. Translated by Thomas Pfau. Albany: State University of New York Press, 1988.
- Novalis. *Fichte Studies/Novalis*. Cambridge, MA: Cambridge University Press, 2003.
- Novalis. *Philosophical Writings* Translated by Margaret Mahony Stoljar. Albany: State University of New York Press, 1997.
- Schlegel, Friedrich von. *Philosophical Fragments/Friedrich Schlegel*. Translated by Peter Firchow. Minneapolis: University of Minnesota Press, 1991.
- Schulte-Sasse, Jochen, ed. *Theory as Practice: A Critical Anthology of Early German Romantic Writings*. Translated by Haynes Horne. Minneapolis: University of Minnesota Press, 1997.
- Wheeler, Kathleen, ed. *German Aesthetic and Literary Criticism: The Romantic Ironists and Goethe*. Cambridge U.K.: New York: Cambridge University Press, 1984.
- Willson, A.L., ed. *German Romantic Criticism*. New York: Continuum, 1982.

*Sebastian Gardner (2005)*

## ROMERO, FRANCISCO (1891–1962)

Francisco Romero, the Argentine philosopher of transcendence, was born in Seville, Spain, but moved to Argentina as a child. After military and literary careers he turned to philosophy, joining the faculty of the University of Buenos Aires in 1928 and of La Plata in 1929. He renounced his academic posts in 1946 in protest against the government of Juan Perón but resumed them in 1955. Because of his conceptual discipline, scope, originality of thought, and limpid clarity of style, Romero is considered one of the ablest and most satisfying of Latin American philosophers.

The idea of transcendence dominates and unifies Romero's metaphysics and theories of knowledge and

values. Transcendence implies at least the diversity achieved by passing beyond a given condition or limit and suggests a universal impetus or agency of such passage, an agency that may be purposive. Opposed to transcendence is immanence, which implies identity and containment within, or return to, a limit. Of the two major forms of transcendence, one is that relation of parts to each other in a structural whole by which novel characteristics emerge that were only latent in the parts considered separately. The other form of transcendence is change and, in particular, evolution in the creative and vitalistic sense of Henri Bergson. Its immanent reduction occurs in the mechanistic evolutionary views of Charles Darwin and Herbert Spencer.

Romero identified reason with immanence; experience, in a broad sense, is related to transcendence. Reason may be either intuitive or discursive. In either case it demands identity and transparency. Identity is found in homogeneity and in permanence; it leads reason to the mechanistic conception of atoms that are similar in kind, endure in time, and are governed by causal laws that presuppose the identification of effects with their causes. Transparency, or clarity, is found in forms emptied of content and in the space in which atoms move and with which they tend to be identified.

Reason is formal only and has no avenue of its own to reality and concrete fact. It is not identical with intelligence, which may criticize it. Where reason fails, experience succeeds. Experience supplies a datum by which knowing must be guided. The objects of experience are not sense data and perceptual objects alone, but also essences and values. In addition, Romero held open the possibility of a metaphysical experience of something ultimate and noumenal but subject to connection with ordinary experience and its phenomenal objects.

Romero divided phenomena into four strata, of which each level is a ground for the next and has greater scope for transcendence than the preceding level. The physical level, that of space and moving atoms, is most pervaded with immanence, but the shift in physical theory from the rigid corpuscle to the *foco activísimo* means a greater emphasis on the role of transcendence even on this level. The vital level is characterized by true duration, a factor of transcendence. The psychical level involves consciousness, which intends, or transcends toward, an object, but there is a countering immanence in the egocentric tendency of the human individual to absorb the object into his own forms and needs. On the spiritual level, the human person, rising above his egocentric needs and attaining a universal subjectivity, contemplates the

object disinterestedly in the sphere of knowing and conducts himself altruistically and with regard to general principles in the sphere of action. On the spiritual level transcendence becomes absolute. The person is transcendence incarnate and unqualified. Each level contains and is supported by transcendence, but each is unique and irreducible.

Romero, proceeding cautiously and with an air of hypothesis, proposed that Arthur Schopenhauer and Bergson were not wrong in positing a metaphysical datum, but that they misconstrued it. Schopenhauer's will and Bergson's vital impulse are forms of transcendence, which is a more general and basic being than either. Romero did not try to sketch the nature of this being, but he appears to have thought of it as a universal impulse at work in every level of phenomenal transcendence, an impetus that is the essence of reality, the source of value, and possibly the spirit's point of flower, which this being intended from the beginning.

*See also* Bergson, Henri; Change; Darwin, Charles Robert; Latin American Philosophy; Nature, Philosophical Ideas of; Schopenhauer, Arthur; Spencer, Herbert.

### Bibliography

#### WORKS BY ROMERO

*Lógica*. Written with E. Pucciarelli. Buenos Aires, 1938.

*Filosofía de la persona*. Buenos Aires, 1944.

*Papeles para una filosofía*. Buenos Aires: Editorial Losada, 1945.

*Teoría del hombre*. Buenos Aires: Editorial Losada, 1952.

Translated by W. F. Cooper as *Theory of Man*. Berkeley: University of California Press, 1964.

*Historia de la filosofía moderna*. Mexico City: Fondo de Cultura Económica, 1959.

Arthur Berndtson (1967)

## RORETZ, KARL

(1881–1967)

Karl Roretz, the Austrian epistemologist, philosopher of culture, and aesthetician, was born at Schloss Breitenreich. He studied law, and later philosophy, at the University of Vienna, receiving his doctorate in 1906 with the dissertation "The Problem of Empathy in Modern Aesthetics." In 1922 Roretz became a *Privatdozent* at the university and taught history of modern philosophy until 1938, when he ceased lecturing after the Nazi takeover of Austria. He resumed lecturing in 1945 and continued until his retire-



ment in 1951.

As an epistemologist, Roretz espoused a “critical positivism,” a philosophy whose foundation is both scientific and, in Immanuel Kant’s sense, criticist. The outstanding features of his thought are critical reflection, skeptical rationality, intellectual honesty, and independence of mind. He rejected dogmatism and unsupported metaphysical speculation. Like Hans Vaihinger, he regarded metaphysical concepts as self-contradictory fictions. Thus, Roretz held, metaphysics lacks any purely logical meaning.

Roretz’s major work, *An den Quellen unseres Denkens* (Vienna, 1937), contains his most acute epistemological analyses. In this monograph he studied “vital concepts,” concepts in whose formation an element of will or an element of value plays a decisive part and whose definition is therefore preceded by a decision. Among such concepts are those of art, of ethics, of popular education, and of the slave trade.

Roretz’s elegant and penetrating psychological analyses of culture and his critical analyses of values deserve particular consideration. The decline of spiritual values, he contended, is due to internal degeneration or disintegration within the person and the society, and only seldom to external pressure. He also studied the genesis and structure of mass psychological phenomena (“mass, illnesses,” *Massenerkrankungen*) in religion, politics, economics, art, fashion, and sports—notably such extremely dangerous religious and other spiritual “epidemics” as belief in vampires and devils, witch-hunting, and racial persecution.

As a philosopher of culture, Roretz felt most akin to Friedrich Nietzsche. Like Nietzsche, he believed in life with a deep conviction. But Roretz’s view of life was Kantian, and the meaning of life for him consisted in working at the problems life poses. He advocated a philosophy that interpreted reality from an aesthetic point of view. Such a philosophy, he held, provides an orientation toward life and the world that is biologically optimal. The world appears, in this view, as a drama without metaphysical supramundane or transmundane galleries to which it must play. Roretz professed a deep joy in the variegated splendor of the world. “The meaning of the world,” he wrote, “is an aesthetic meaning.”

In his studies of what he called intellectual-aesthetic values—aesthetic effects bound up with specific achievements of thought, as in mathematics, strategy, or chess—Roretz made important contributions to aesthetics itself.

His interest in ethical problems was equally great. A convinced humanist and democrat, he supported the Ethical Culture movement and strove for a secular ethics independent of any metaphysical or religious assumptions.

**See also** Aesthetics, History of; Culture and Civilization; Epistemology; Kant, Immanuel; Nietzsche, Friedrich; Value and Valuation.

### **Bibliography**

Additional works by Roretz include *Zur Analyse von Kants Philosophie des Organischen* (Vienna, 1922); *Religiöse Epidemien* (Munich, 1925); *Die Metaphysik—eine Fiktion* (Vienna, 1927); and *Ziele und Wege philosophischen Denkens* (Wien: F. Deuticke, 1976).

Also see Franz Austeda, *Dem österreichischen Philosophen K. Roretz zum achtzigsten Geburtstag* (Vienna, 1961), which contains a complete bibliography.

**Franz Austeda (1967)**

*Translated by Albert E. Blumberg*

*Bibliography updated by Michael Farmer (2005)*

## RORTY, RICHARD

(1931–)

An American philosopher and pragmatist, Rorty is among the most widely discussed and controversial philosophers at the turn of the twenty-first century. A New Yorker by birth, Richard Rorty was educated at the University of Chicago (1946–1952) and at Yale (1952–1956) where he received his doctorate in philosophy. After brief flirtations with Platonism and the work of A. N. Whitehead, Rorty’s more mature interests began to form at the end of his military service in 1958, at which point he began serious study of the philosophers who would later number among his chief influences: Wilfrid Sellars, Ludwig Wittgenstein, Martin Heidegger, John Dewey, and W.V. O. Quine.

### EARLY PERIOD

Rorty’s early work in analytic philosophy, sometimes thought to represent a completely distinct period, is in fact touched by two themes that resurface throughout his career. The first theme is anti-Cartesianism about the mind and knowledge. In a series of papers written during the 1960s Rorty was the first to develop a subsequently contentious theory in the philosophy of mind—eliminative materialism, which holds that the mind and mental

states are theoretical, and hence dispensable, constructions.

The second theme is an abiding concern with the function and importance of philosophy. Again, this theme appears early on, particularly in Rorty's 1967 introduction to *The Linguistic Turn*, a collection of essays on analytic philosophy of language. In his introduction, Rorty praised analytic philosophy for knocking the entire philosophical tradition on its heels—a sentiment that he would later characterize as naïve. In subsequent work, Rorty came to believe that mainstream Anglo-American philosophy of language makes many of the same mistakes as the intellectual traditions he had earlier taken it to supplant.

### ANTIREPRESENTATIONALISM

This latter sentiment first emerges in Rorty's seminal book, *Philosophy and The Mirror of Nature* (hereafter: PMN). Originally published in 1979, and at once hailed and denounced as a critique of analytic philosophy, the book brings together Rorty's hostility to Cartesianism with a positive vision about the nature and limits of philosophy.

At its core, PMN is a sustained attack on representationalism. According to representationalism—which Rorty argues is a largely unquestioned assumption of Cartesian and Kantian philosophy—the mind is a device for representing the world and knowledge is accurate representation. Rorty holds the representationalism responsible for two major philosophical mistakes: a false dualism of mind and body, and a bankrupt foundationalist picture of knowledge, which holds that all knowledge must sit on the foundation stones of intrinsically accurate privileged representations. Moreover, Rorty takes representationalism to paint a misleading picture of philosophy's importance—as a master discipline that judges whether the claims of science, morality, or art can represent reality.

In Rorty's view, twentieth-century linguistic philosophy continued to assume representationalism (and its mistakes) in a linguistic guise—an opinion he sees as shared by philosophers as diverse as Dewey, Quine, Sellars, Wittgenstein, and Donald Davidson. Following Quine and Wittgenstein, for example, he argues that the notions of meaning and analyticity are mere linguistic shadows of the privileged mental representations of the early moderns. And with Sellars, he rejects “the given,” or theoretically innocent sense-experience, as a myth. His moral: language, like the mind, should not be understood as a device for representing a ready-made world.

In opposition to representationalism, Rorty suggested in PMN that we should adopt what he called *epistemological behaviorism*, and explain epistemic rationality and justification in terms of what our society will let us say, rather than the other way around. The thought is that there is no mystery about how the mind represents the world. The very idea of such representation is a fable; what claims we accept as knowledge depends not on how well they mirror the world but on how well they hang together with what else we already accept. Accordingly, Rorty concluded, philosophy has nothing distinctive to offer about knowledge. To learn why we accept what we accept we must turn to biology, psychology, and sociology; Charles Darwin will have more to teach us about the mind than René Descartes. The philosopher's role was instead therapeutic—to cure of us intellectual maladies—and revisionary so as to convince us to engage in new forms of conversations.

### PRAGMATISM AND TRUTH

Rorty has continually emphasized that his view is a form of pragmatism—particularly the pragmatism of Dewey. And much like the classical pragmatists before him, he sees his debate with the representationalist as coming down to a debate over how to understand truth.

Yet Rorty's own views on truth shifted in subtle ways over the course of his career. He always rejected the correspondence theory of truth, according to which a statement is true just when it corresponds to the facts. But in his earlier work, Rorty was tempted to follow the classical pragmatists and define truth in terms of justification or warranted assertibility. Truth, on his version of the view, simply is what we are justified in believing in light of our cultural practices. But in later works, Rorty has come to see this position as another misguided attempt to uncover the secret nature of truth. The contemporary pragmatist, Rorty argues, should instead simply reject the idea that truth has any nature at all. Truth is not the sort of thing that can be defined—not because its nature is mysterious or ineffable, but because there is nothing general and informative one can say about what is in common between “Snow is white,” “Two and two are four” and “Democracy is a better form of government than tyranny.” There simply is no metaphysically substantive property of truth that some propositions have and others lack.

Rorty argues that adopting this attitude toward truth has several important consequences. In his later work, for example, he has particularly emphasized that, for the pragmatist, truth is not a goal of inquiry. According to

Rorty, something can only be a goal if we can recognize when we have reached it. But whenever we check to see whether our beliefs are true, we can only discover whether they are justified or unjustified. Thus we should give up on the idea that we are aiming at truth; instead, Rorty says, echoing Davidson, we should see ourselves as aiming only at honest justification. And for Rorty, justification is a practical matter—what beliefs we find justified depends on whether we can use them in achieving the aims of our culture. Nonetheless, truth is not reducible to what our immediate community finds useful because one important function of the word *true* in our language is to remind us that what may be practically justified to some audiences may not be justified to all.

Rorty's views on truth have drawn considerable criticism. He is often derided as advocating a naïve form of cultural relativism. But Rorty insists that it is as misleading to describe him as a relativist as it would be to describe him as a realist. In Rorty's eyes, the realist and the relativist commit linked sins: the realist by taking the world to be ready-made, the relativist by thinking it is made by us. From the Rortyan perspective, we should instead take truth making—whether understood in a realist or relativist fashion—as simply a metaphor that should be given up. Consequently, Rorty might be better described as advocating a form of quietism about metaphysics and epistemology.

## DEMOCRACY AND PHILOSOPHY

Rorty takes the failure of representationalism as linked with the failure of another enlightenment project: the project of grounding our political ideals in a common human nature. For Rorty, democratic, liberal government is a great achievement, but it is not an achievement whose value can be given a philosophical justification. Rather than trying to justify liberal democratic ideals philosophically, we should instead seek to ground our philosophical ideals in our democratic values. Thus we should stop searching for objective foundations and instead aim for solidarity with our fellows.

Rorty describes his positive political position as *liberal irony*. It is liberal because it takes self-creation and freedom as central values. Individuals should be free from suffering and cruelty, but also free to create and live their own vision of the good life. But the Rortyan liberal also takes an ironic stance toward his own liberal commitments. He realizes that his values are contingent reflections of his own time and place, and not reflections of the values of the world itself. To those critics who protest that this position is too weak to offer sufficient defense against

the tyrant, Rorty responds that philosophy is of no use against tyranny anyway, and that those who believe that all is lost without appeal to the world's own true values are much like those nineteenth-century intellectuals who believed that without God, everything was permitted.

Rorty has sometimes been charged with no longer doing philosophy. And that charge is fair if one takes philosophy to be in the job of representing the world as it is in itself. But Rorty's own views encourage a different view of philosophy, according to which the job of the philosophy is not so much to discover the world's own language as he sometimes put it, but to invent new vocabularies and means of description. In this sense, Rorty's stance toward philosophy is Marxian: The goal of the philosopher is not to map the landscape as it is, but change how we see the world; to paint new landscapes, new pictures.

**See also** Epistemology; Philosophy of Mind; Pragmatism; Social and Political Philosophy.

## Bibliography

### PRIMARY WORKS

- Philosophy and The Mirror of Nature*. Princeton, NJ: Princeton University Press, 1979.
- Consequences of Pragmatism*. Minneapolis: University of Minnesota Press, 1982.
- Contingency, Irony, and Solidarity*. Cambridge, MA: Cambridge University Press, 1988.
- Objectivity, Relativism, and Truth: Philosophical Papers I*. Cambridge, MA: Cambridge University Press, 1991.
- Essays on Heidegger and Others: Philosophical Papers II*. Cambridge, MA: Cambridge University Press, 1991.
- The Linguistic Turn*. 2nd ed. Chicago: University of Chicago Press, 1992.
- Truth and Progress: Philosophical Papers III*. Cambridge, MA: Cambridge University Press, 1998.

### SECONDARY WORKS

- Brandom, Robert. *Rorty and His Critics*. Oxford: Blackwell Publishers, 2000.

*Michael P. Lynch (2005)*

## ROSCELIN (1050–1120)

Despite much scholarly effort, little is known about Roscelin of Compiègne. The only work that we can safely attribute to him is a letter he sent to Peter Abelard around 1119–1120. In this ill-tempered piece of writing Roscelin sets out to distinguish his position on the Trinity from

that which Abelard was developing in his *Theologia Summi Boni*.

The problem for Roscelin and Abelard is to give an account of the distinction between the persons of the Trinity compatible with the unity of God. Roscelin notes that he has to navigate here between two heresies: Sabellianism, requiring such a unity in the singular substance of God that the distinction between the persons can be only verbal, and Arianism, which distinguishes the persons as greater and lesser so as to constitute three distinct gods. Roscelin, in effect, accused his former student of Sabellianism, and so contributed to Abelard's being called before the Council of Soissons in 1121 and required to burn a copy of his *Theologia Summi Boni* and confirm his orthodoxy.

Roscelin's position is that the names *Father*, *Son*, and *Holy Spirit* pick out distinct items but that these exist in God with the unity of likeness and equality. Unfortunately he invokes authority rather than arguing for his theory. It has recently been suggested by author Constant Mews, however, that Roscelin depends on an account of the semantics of names developed in contemporary writing on grammar.

Our remaining information about Roscelin comes from two unsympathetic reporters. Abelard, in a letter written around 1119, complains about Roscelin's attack on him and recalls that long before, at the Council of Soissons in 1092, Roscelin had been charged with tritheism. From Saint Anselm we have two letters from 1090–1092, and the treatise *De incarnatione verbi*, written after the Council, in which Roscelin is said to have maintained that the persons of the Trinity are as separate from one another as three angels, or three souls. This is certainly not what he claims in his letter to Abelard, which may thus represent a refinement of his theory in response to Anselm's objections. In *De incarnatione verbi* Anselm characterizes Roscelin as a heretic in logic who holds that universal substances are nothing more than "puffs of air made with the voice," who cannot distinguish a body from its color, or a soul from its wisdom, and cannot understand how human beings are one in species.

In the middle of the twelfth century Roscelin was said to have been the first to have upheld the doctrine of words (*sententia vocum*). From the information given, however, it is impossible to recover anything of his theory. We are told by John of Salisbury that he held that utterances themselves were genera and species. This is the position advocated in the *Dialectica* of Garlandus Compotista, written around 1115, which may provide our best

guide to the views of Roscelin and those referred to at the time as the *Vocales*.

Appealing to Garlandus, Abelard's early writings, and various other texts, author John Marenbon has argued that vocalism in general and perhaps Roscelin's views in particular, developed out of what he calls the *in voce* reading of Porphyry's *Isagoge* and Aristotle's *Categories*. In this exegetical procedure, he suggests, theoretical commitment was suppressed in favor of reading the texts as simply about the relations between words.

Abelard confirms in his letter that Roscelin held that universals are in some sense words, and parodies him by saying he would have to read Scripture as claiming that Christ ate the expression *broiled fish* rather than the fish itself. It is unfortunately impossible to tell whether Abelard is constructing or reporting an argument when he reports, in his *Dialectica*, that as well as holding that species are words, Roscelin claimed that things do not have parts—so a wall is not part of a house. Perhaps what was really at issue were the questions that seem to have exercised Roscelin throughout his career: What counts as a thing and what is the nature of unity?

**See also** Abelard, Peter; Anselm, St.; Aristotle; Arius and Arianism; John of Salisbury; Medieval Philosophy; Porphyry; Universals, A Historical Survey.

### **Bibliography**

- Marenbon, John. "Life, Milieu, and Intellectual Contexts." In *The Cambridge Companion to Abelard*. Cambridge, U.K.: Cambridge University Press, 2004.
- Mews, Constant. *Reason and Belief in the Age of Roscelin and Abelard*. Aldershot U.K.: Ashgate, 2002.

*Christopher J. Martin (2005)*

## ROSENKRANZ, JOHANN KARL FRIEDRICH (1805–1879)

Johann Karl Friedrich Rosenkranz, the German Hegelian philosopher, was born in Magdeburg. He entered the University of Berlin in 1824. Although he was to become G. W. F. Hegel's most devoted disciple, Rosenkranz was first drawn to Friedrich Schleiermacher; he heard only an occasional lecture by Hegel and was unimpressed. He began reading Hegel as a student at Halle in 1826 and the following year came under the influence of Karl Daub (1765–1836), a Hegelian theologian at Heidelberg. As a *Privatdozent* and extraordinary professor at Halle,

Rosenkranz participated actively in the Hegelian circle there. Called to Berlin, he struck up a friendship with Hegel and joined his birthday celebration a few weeks before Hegel died of cholera in 1831. Rosenkranz himself was stricken almost fatally with the disease, reflecting, as he later reported, that this was carrying discipleship entirely too far. In 1833 he succeeded Johann Friedrich Herbart as professor of philosophy at the University of Königsberg, where he remained until his death except for a brief political career in Berlin during the revolutionary crisis of 1848/1849.

Rosenkranz wrote over forty substantial works, on systematic philosophy, aesthetics, theology, logic, psychology, literary history, pedagogics, philosophical history and biography, and political and social theory. He also composed poetry and contributed articles on current issues to the newspapers.

Rosenkranz defended the Hegelian system as the authentic expression of the German spirit and the fulfillment of German philosophy. He attacked the “one-sidedness” of the Hegelian left-wing and denied that there was any irreconcilable conflict between Hegel and other major German thinkers, such as Schleiermacher and Immanuel Kant. Other Hegelians charged that Rosenkranz had interpreted Hegel in a Kantian way, maintaining the duality between thought and being and between the ideal and the actual. Certainly in his view the ideal was always in tension with existing conditions, although it constituted their *telos* and guiding norm. In practice, for example, he held that the church should be independent of the state; because Christianity embodies the highest ideal, the church must be free to hold before the culture its most ideal possibilities. He argued on similar grounds for the freedom of the university from political control.

Underlying religious, political, and intellectual life alike, however, was the *Volksgeist* (“spirit of a people”), interpreted more romantically than in Hegel. It is not the result of the cultural process but the distinctive psychic root of a particular people that gives the people unity as a nation and seeks expression in a total cultural life. A people is free to the extent that it fully embodies this spirit; genuine “public opinion” is the self-understanding of a free people. As a consequence, although Rosenkranz gave humankind precedence over the nation in principle and affirmed the Kantian vision of universal peace, he opposed the supranationalism of the left-wing Hegelians; moreover, he regarded their revolutionary aims as empty abstractions, without relevance to “realities” or to the concrete aspirations of any people, and productive only

of despotism. He advocated German unification, under a constitutional monarchy and through Prussian initiative, but only under a constitution that would express the German spirit. Although he vigorously opposed revolutionary change in the Prussian form of government, he just as vigorously, and at personal risk, attacked the repressive policies of its administration. For example, he defended the freedom of the press as the organ of “public opinion”; the local press, in turn, hailed him as “the most popular and liberal man in Königsberg.”

**See also** Hegel, Georg Wilhelm Friedrich; Hegelianism; Herbart, Johann Friedrich; Kant, Immanuel; Schleiermacher, Friedrich Daniel Ernst.

### Bibliography

#### WORKS BY ROSENKRANZ

- Encyclopädie der theologischen Wissenschaften*. Halle, 1831.
- Kritik der Schleiermacherschen Glaubenslehre*. Königsberg, 1836.
- Psychologie*. Königsberg: Gebrüder Bornträger, 1837.
- Studien*. 5 vols. Berlin and Leipzig: Jonas Verlagsbuchhandlung, 1839–1848.
- Kritische Erläuterungen des Hegelschen Systems*. Leipzig: Gebrüder Bornträger, 1840.
- Geschichte der Kantischen Philosophie*. Leipzig, 1840. Vol. XII of Kant's *Werke*, edited by Rosenkranz and Schubert.
- Schelling*. Danzig: F.S. Gerhard, 1843.
- G. W. F. Hegels Leben*. Berlin: Duncker and Humblot, 1844. Supplement to Hegel's *Werken*.
- System der Wissenschaft: Ein philosophisches Enchiridion*. Königsberg: Gebrüder Bornträger, 1850.
- Asthetik des Hässlichen*. Königsberg: Gebrüder Bornträger, 1853.
- Wissenschaft der logischen Idee*. 2 vols. Königsberg: Gebrüder Bornträger, 1858–1859.
- Hegel als deutscher Nationalphilosoph*. Leipzig: Duncker and Humblot, 1870.
- Neue Studien*. 4 vols. Leipzig, 1875–1878.
- Politische Briefe und Aufsätze 1848–1856*, edited by Paul Herre. Leipzig: Dieterich, 1919.
- Die Hegelsche Rechte*, edited by Hermann Lübke. Stuttgart and Bad Cannstatt, 1962. Contains several essays, mostly political, and complete bibliography.

#### WORKS ON ROSENKRANZ

- Esau, Lotte. *Karl Rosenkranz als Politiker*. Halle: M. Niemeyer, 1935.
- Jonas, Richard. *Karl Rosenkranz*. Stuttgart: Weicher, 1906.
- Metzke, Erwin. *Karl Rosenkranz und Hegel*. Leipzig: Heims, 1929.

Stephen D. Crites (1967)

## ROSENZWEIG, FRANZ

(1886–1929)

Franz Rosenzweig, the religious existentialist, was born in Cassel, Germany. From 1905 to 1912 he studied natural sciences, modern history (under Friedrich Meinecke), and philosophy (under Heinrich Rickert) at the universities of Göttingen, Munich, Freiburg, and Berlin. At Berlin he earned a doctor of philosophy degree in 1912 with a dissertation on G. W. F. Hegel's political doctrines; later, he expanded this study. In the fall of 1913, after a spiritual crisis, he turned to religious, especially Judaic, philosophy. In 1918–1919 he wrote *Der Stern der Erlösung* (*The Star of Redemption*), a three-part religio-philosophical system; in 1920 he founded the Freies Jüdisches Lehrhaus (Independent House of Judaic Studies) in Frankfurt. Two years later he was appointed lecturer for Jewish religious philosophy and ethics at the University of Frankfurt, but the onset of progressive paralysis prevented him from accepting the appointment. Despite his affliction, he continued his scholarly work until his death in Frankfurt.

*Hegel und der Staat* (Hegel and the state), completed in 1914, for which Rosenzweig used both published and unpublished materials, analyzes the development of Hegel's concept of the state and its place in the philosopher's system. For Rosenzweig, the reasons motivating the successive changes in Hegel's political theories are to be found in the philosopher's intellectual progression.

In "Das älteste Systemprogramm des deutschen Idealismus" (The earliest systematic program of German idealism; written in 1914), Rosenzweig established that young Friedrich von Schelling was the author of a treatise preserved in Hegel's handwriting. This treatise is Schelling's sole attempt at formulating a unified system, a feat most perfectly realized by Hegel.

Rosenzweig's own philosophy may be defined as religious existentialism. *The Star of Redemption* begins with a critique of the Western philosophic tradition and, especially, of Hegel. Rosenzweig rejected as contrary to experience the attempt to reduce to one basic essence the three elements of reality: God, the world, and man.

In German idealism it is human consciousness and thought from which both God and world are deduced. In addition, consciousness is understood as "consciousness in general," which reduces to insignificance the individual being and his separate consciousness. But thought, Rosenzweig argued, is only one of the components of existence; it does not precede existence. The significance of the individual man stems from his being alive; he is more than a part of nature or the world. In this affirma-

tion of the concrete person in his particularity Rosenzweig resumed the anti-Hegelian revolt of Arthur Schopenhauer, Ludwig Feuerbach, Søren Kierkegaard, and Friedrich Nietzsche, with its concern for the individual. The experience (*Erfahrung*) of the thinker, intent upon the value and significance of things, must guide him in confronting existence. Experience offers knowledge of God, the world, and man.

Under the influence of the later Schelling, and, to a certain degree, of Hermann Cohen, Rosenzweig links his theory of experience with a theory of conceptual construction; this linkage helps him to discover the interrelationship and interaction of the elements of God, world, and man. By way of an intricate logical construct he arrives at the following statement of relationships in terminology borrowed from theology: creation denotes the action of God upon the world; revelation, the encounter of God and man; and redemption, the relation of man to the world.

In pagan imagery God, the world, and man are separated and independent of each other. The hero of Greek tragedy is isolated from men and alien to the gods; the plastic cosmos is unrelated to man and the gods, who, in turn, have no concern for the world or man. Only biblical religion teaches the interaction of the elements of reality; in this concept, added to what he calls experience, lie the roots of Rosenzweig's existentialism. According to this view, creation is the process through which God, hitherto hidden in the mythical beyond, appears to give the world reality. But creation implies transitoriness, finiteness, death; the process of creation is renewed and perfected in revelation, through which God, in his love, turns to man; the experience of this love evokes in man the consciousness of being a self and accords man reality. Now his original isolation and dumbness are overcome; his response to God's love is his own love. Man translates his love for God into love for his "neighbor," and by so doing participates in leading the world toward redemption. Through the deeds of love the temporality of life and the finality of death are overcome. Ultimate redemption is anticipated, and a sense of eternity in time experienced, primarily in the rhythm of the days that constitute the sacred calendar in the religions based on revelation, Judaism and Christianity. Both these religions represent, under the aspect of faith, authentic, though different, manifestations of reality, and both are concerned with the existential situation of individual man.

The ideal representative of the "new thinking," as Rosenzweig called his view, is a philosopher-theologian who, while maintaining scholarly objectivity, accepts the

subjective, unique self as the new point of departure. The new theology should be existentially orientated, and theological problems should be translated into human terms. In contradistinction to abstract, timeless, purely logical, solitary thinking, the new existential thinking is “grammatical”: Human language, the word, the name, dialogue, are keys to the understanding of reality; the speaking thinker thinks *for* someone and speaks *to* someone. In such language-bound thinking, utmost importance is accorded to time; past, present, and future are actively involved in the process of thought, a notion found also in Martin Heidegger’s philosophy.

In the Judaic field, Rosenzweig advocated a reevaluation of the thought of classical Judaism. With Martin Buber he undertook to translate the Old Testament, faithfully transposing into German the style of the original.

**See also** Buber, Martin; Cohen, Hermann; Consciousness; Existentialism; Feuerbach, Ludwig Andreas; Hegel, Georg Wilhelm Friedrich; Heidegger, Martin; Idealism; Jewish Philosophy; Kierkegaard, Søren Aabye; Meinecke, Friedrich; Nietzsche, Friedrich; Revelation; Rickert, Heinrich; Schelling, Friedrich Wilhelm Joseph von; Schopenhauer, Arthur.

### Bibliography

#### WORKS BY ROSENZWEIG

- Hegel and der Staat*. 2 vols. Berlin: Oldenbourg, 1920.
- Der Stern der Erlösung*. Frankfurt am Main: Kauffmann, 1921; 3rd ed., Heidelberg, 1954. An English version is *The Star of Redemption*, translated by William W. Hallo. New York: Holt, Rinehart and Winston, 1971.
- Briefe*, edited by Ernst Simon and Edith Rosenzweig. Berlin: Schocken, 1935. A selection of letters.
- Kleinere Schriften*. Berlin: Schocken, 1937. A collection of papers, including “Das älteste Systemprogramm des deutschen Idealismus” and “Das neue Denken” (1925), a discursive “epilogue” to *Der Stern*.
- Das Büchlein vom gesunden und kranken Menschenverstand*. Düsseldorf, 1964. Written in 1921. Includes a popular presentation of the chief theme of *Der Stern*. An English version is *Understanding the Sick and the Healthy*, edited by N. N. Glatzer. New York: Noonday Press, 1953.

#### WORKS ON ROSENZWEIG

- Altmann, Alexander. “Franz Rosenzweig on History.” In *Between East and West*. London, 1958.
- Buber, Martin. “Franz Rosenzweig.” *Kant-Studien* 35 (4) (1930): 517–522.
- Freund, Else. *Die Existenzphilosophie Franz Rosenzweigs*. 2nd ed. Hamburg, 1959.
- Glatzer, Nahum N. *Franz Rosenzweig: His Life and Thought*. New York: Farrar, Straus and Young, 1953.

Guttmann, Julius. “Franz Rosenzweig.” In *Philosophies of Judaism*. Translated by D. Silverman, 367–398. New York, 1964.

Löwith, Karl. “M. Heidegger and F. Rosenzweig, or Temporality and Eternity.” *Philosophy and Phenomenological Research* 3 (1) (1942): 53–77.

*Nahum Norbert Glatzer (1967)*

## ROSENZWEIG, FRANZ [ADDENDUM]

Rosenzweig drew heavily from the lectures of his teacher in modern Jewish philosophy, Hermann Cohen, to construct his own highly original reevaluation of the thought of classical Judaism on the model of Judah Halevi’s philosophy and poetry within the framework of the post-Hegelian, post-rationalist, German Romantic philosophy of the early twentieth century. Rosenzweig expressed his Jewish thought through many forms, including new German translations of the Hebrew Scriptures, essays on Jewish education, and his personalized administration of a nonaccredited school for Jewish studies at the University of Frankfurt. No Jewish theologian has had a more lasting impact on the subsequent development of Jewish philosophy than has Franz Rosenzweig. It is not an exaggeration to say that with very few exceptions every important Jewish religious thinker in the second half of the twentieth century was either his student or a student of his students in the United States, in Israel, and in western Europe.

**See also** Cohen, Hermann; German Philosophy; Halevi, Yehuda; Jewish Philosophy; Rosenzweig, Franz.

### Bibliography

#### PRIMARY WORKS

- Buber, Martin, and Franz Rosenzweig. *Schrift und ihre Verdeutschung* [Scripture and Translation]. Translated by Lawrence Rosenwald with Everett Fox. Bloomington, IN: Indiana University Press, 1984.
- Rosenzweig, Franz. *Cultural Writings of Franz Rosenzweig*. Translated by Barbara E. Galli. Syracuse, NY: Syracuse University Press, 2000.
- Rosenzweig, Franz. *Die “Gritli” – Briefe an Margrit Rosenstock-Huessy*. Tübingen, Germany: Verlag Bilam, 2002.
- Rosenzweig, Franz. *Der Mensch und sein Werk: Gesammelte Schriften*. The Hague, Netherlands: Martinus Nijhoff, 1976.
- Rosenzweig, Franz. *Ninety-Two Poems and Hymns of Yehuda Halevi*, edited by Richard A. Cohen. Translated by T. Kovach, E. Jospe, and G.G. Schmidt. Ithaca, NY: State University of New York Press, 1999.

- Rosenzweig, Franz. *On Jewish Learning*. New York: Schocken Books, 1995.
- Rosenzweig, Franz. *Der Stern der Erlösung*. The Hague, Netherlands: Martinus Nijhoff, 1976. Translated into English by William Hallo as *The Star of Redemption*. Notre Dame, IN: Notre Dame Press, 1985. Translated into Hebrew by Yehosha Amir as *Kokhav Ha-Geulah*. Jerusalem: Mosad Bialik, 1970.
- Rosenzweig, Franz. *Understanding the Sick and the Healthy: A View of World, Man, and God*. Cambridge, MA: Harvard University Press, 1999.

SECONDARY WORKS

- Anckaert, L., and Casper, B. *Franz Rosenzweig: A Primary and Secondary Bibliography*. Leuven, Belgium: Bibliotheek Van De Faculteit Der Godgeleerdheid Van De K. U. Leuven, 1990.
- Batnitzky, Leora. *Idolatry and Representation: The Philosophy of Franz Rosenzweig Reconsidered*. Princeton, NJ: Princeton University Press, 2000.
- Cohen, Richard A. *Elevations: The Height of the Good in Rosenzweig and Levinas*. Chicago: University of Chicago Press, 1994.
- Freund, Else Rahel. *Franz Rosenzweig's Philosophy of Existence: An Analysis of the Star of Redemption*. The Hague, Netherlands: Martinus Nijhoff, 1979.
- Funkenstein, Amos. "The Genesis of Rosenzweig's 'Stern der Erlösung': 'Urformel' and 'Urzelle.'" *Jahrbuch des Instituts für Deutsche Geschichte* Beiheft 4 (April, 1982).
- Galli, Barbara Ellen. *Franz Rosenzweig and Jehuda Halevi*. Montreal: McGill-Queen's University Press, 1995.
- Gibbs, Robert. *Correlations in Rosenzweig and Levinas*. Princeton, NJ: Princeton University Press, 1992.
- Glatzer, Nahum Norbert, ed. *Franz Rosenzweig: His Life and Thought*. New York: Farrar, Straus & Young, 1953.
- Gordon, Peter Eli. *Rosenzweig and Heidegger: Between Judaism and German Philosophy*. Los Angeles: University of California Press, 2003.
- Mendes-Flohr, Paul R. ed. *The Philosophy of Franz Rosenzweig*. Hanover, NH: University Press of New England, 1988.
- Miller, Ronald H. *Dialogue and Disagreement: Franz Rosenzweig's Relevance to Contemporary Jewish-Christian Understanding*. Lanham, MD: University Press of America, 1989.
- Mosès, Stéphane. *System and Revelation: The Philosophy of Franz Rosenzweig*. Translated by Catherine Tihanyi. Detroit: Wayne State University Press, 1992.
- Oppenheim, Michael D. *What Does Revelation Mean for the Modern Jew?: Rosenzweig, Buber, Fackenheim*. Lewiston/Queenston, U.K.: The Edwin Mellon Press, 1985.
- Rotenstreich, Nathan. *Jews and German Philosophy: The Polemics of Emancipation*. New York: Schocken, 1984.
- Rubinstein, Ernest. *An Episode of Jewish Romanticism: Franz Rosenzweig's The Star of Redemption*. Albany, NY: SUNY, 1999.
- Samuelson, Norbert M. "The Philosophy of Rosenzweig." Chaps. 27–29 in *Jewish Philosophy: An Historical Introduction*. London: Continuum, 2003.
- Samuelson, Norbert M. *User's Guide to Franz Rosenzweig's Star of Redemption*. Richmond, U.K.: Curzon, 1999.

- Santner, Eric L. *On the Psychotheology of Everyday Life: Reflections on Freud and Rosenzweig*. Chicago: University of Chicago Press, 2001.
- Turner, Joseph. *Faith and Humanism: A Study in Franz Rosenzweig's Religious Philosophy*. In Hebrew. Tel Aviv, Israel: Hakibbutz Hameuchad, 2001.
- Vogel, Manfred. *Rosenzweig on Profane/Secular History*. Atlanta: Scholars Press, 1996.

Norbert Samuelson (2005)

ROSMINI-SERBATI,  
ANTONIO  
(1797–1855)

The Italian philosopher, educator, and statesman Antonio Rosmini-Serbati was born in Rovereto, then part of the Austrian Tyrol. The families of both his parents held patents of nobility under the Holy Roman Empire. A private education begun at an early age and directed to the priesthood established a firm foundation for his later work. Finding Austrian rule oppressive, Rosmini moved to the freer region of Piedmont. He started his career by founding the Institute of Charity, devoted to education and missions. He began to publish prolifically in philosophy, literature, and pedagogy. In politics he became an active exponent of the principles of Neo-Guelphism and reached the peak of his public career as counselor to Pius IX during the period from 1848 to 1853; at the end of this period, more conservative forces came to power. Retiring to private life, Rosmini continued his writing and assumed the direction of his institute. The present article restricts itself to Rosmini's philosophical work.

Although developed in a large number of works, Rosmini's philosophical thought presents a high degree of unity. This unity has two sources: the historical and apologetic intentions that sustain it and the internal development of certain germinal ideas. Rosmini's overt intention was to create a Christian-Catholic apologetics that would meet the demands of modern philosophical thought while remaining faithful to the core of traditional Christian philosophy. Since Augustinian and Thomistic realism predominated in Christian philosophy, Rosmini endeavored to anchor his thought in that tradition, exhibiting an affinity to the Augustinian strain. At the same time, he sought to meet the demands of rationalism and empiricism, and especially of the Kantian attempt at a resolution of the tension between the two. The effort to meet these conditions imparted to Rosmini's thought a high degree of complexity and sophistication.



The point of departure of the Rosminian system is his *Nuovo saggio sull'origine delle idee* (1830), a work of elaborate synthesis. The controlling principle of the synthesis is basically Augustinian, but the work develops around three centers: the idea of being, intellectual perception, and the origin of ideas.

### THE IDEA OF BEING

Following Immanuel Kant, Rosmini accepted a dual order of a posteriori and a priori in the process of knowledge and identified the ground of science with a priori principles of knowledge. Whereas Kant distinguished diverse orders or forms of a priori synthesis, Rosmini reduced that plurality to a single form, the idea of being. Only the idea of being can be thought without reference to any other idea, and only that idea is thought, at least implicitly, in the thinking of any other idea. The idea of being is not the product of the subject, whether empirical or transcendental; it is a datum offered immediately by God to the intelligent subject; it is, moreover, ontologically and functionally constitutive of that subjectivity. The idea of being is both a category and a transcendental operation. It is a category, for the subject knows through the process of the existential judgment, in which being as given in the idea of being is predicated of things. This judgment establishes the subsistence of the object as present and known in the judgment.

As a category, the idea of being is the irreducible "other" to any specific content of thought or knowledge. It must also either be a product of the empirical subject or be truly objective. In the first case, the idea of being would be subjective and would render all knowledge subjective; in the second, its objectivity would seem to require the postulation of a "transcendental" subject. Rosmini accepted neither horn of this ostensible dilemma. He held that the human subject is empirical but also capable of a transcendental operation by which it can secure universal and necessary knowledge. It performs this operation through the idea of being; more accurately, this operation is one with the idea of being. As a transcendental operation, the idea of being constitutes the knowing subject ontologically and existentially; it secures the realm of universal and necessary knowledge. Finally, it is transcendent, for it is not the product of the subject, whether empirical or transcendental, but a datum that must be referred to the action of God. It is this last point that relates Rosmini's view to that of Augustine.

### INTELLECTUAL PERCEPTION

Although no knowledge is possible except through the idea of being, that idea does not suffice for the effective knowledge of the actual world of determinate forms of subsistence. This world can be known only if sensation has entrance into the realm of the idea of being and vice versa. Sensation is the vehicle of the multiple forms of determinate subsistence of the real world, but it does not present them as being; for them to be presented as being, sensation must be infused by the idea of being. This infusion is achieved concretely in an operation that Rosmini called intellectual perception.

Intellectual perception is rooted in man's fundamental constitution, for he is both sentient and intelligent. Every concrete act of knowing is structured by sensation and intelligence, related in a radical unity. There is neither pure sensation nor pure intellection, or intellectual vision. Intellectual perception, in which these pure elements occur in vital union, places man in authentic contact with the concrete real world. This operation is perception because by it the subject sensibly lays hold of reality, which actually stands before it, as subsistent. It is intellectual because the sensible perception evokes in the indeterminate being, which is already present to the subject in the idea of being, determinations by which the ideas of particular things arise. Intellectual perception is not, manifestly, the synthesis of two antecedently existing elements; it is the complex term of a complex, concrete operation, rooted in the fact that man is a complex principle and subject, both intelligent and sensitive.

### ORIGIN OF IDEAS

On the basis of the foregoing points, Rosmini addressed the problem of the origin of ideas. Ideas, except the idea of being, arise through the process of abstraction. Empiricists and sensationists confuse intellectual perception with sensation when they speak of the formation of ideas out of the elements of sensation through abstraction and reflection. The act of reflection is not performed on the simple sense datum but upon objects already known and present through intellectual perception. By noting certain characteristics and averting attention from others, abstraction forms ideas of various degree up to the most general. The idea of being is alone excluded from this account; for it is the presupposition, not the result, of intellectual perception.

### SUBJECTIVE REALISM

Rosmini proceeded in *Psicologia* (1850) to consider the subject, which is the locus of the process of knowledge.

Here again his doctrine reflects his concern to meet both empirical and idealist claims by passing beyond them. He refused to resolve the subject into the transcendental process, as he claimed idealists did, or into the process of sensation, as he said empiricists did. Instead, he offered a “subjective realism” or, better, a “realism of the subject.” Its basis is the theory of the “fundamental sentiment,” the immediate analogue for which is intellectual perception. The soul, while retaining its classical status as the active principle of vital operations and psychic phenomena, takes on a new dimension; it is the substance-sentiment, the intuitive sense of immanent being that generates subsistence. The reality of the subject is constituted by this immediate, nonobjective, and synthetic sense of self, which draws into a subsistent unity all aspects—sensitive, intelligent, and volitional—of the subject’s complex life. This fundamental sentiment is the first and the continuous experience that man has of himself. It always involves, moreover, a relationship to a corporeal term, the body. This specific aspect of the fundamental sentiment, the corporeal sentiment, is characteristic of human nature. By it Rosmini justified the classical doctrine of man’s composition out of body and soul. All other sensations are accidental to this fundamental sentiment; it is primitive and incommunicable and constitutes the subject in its unity and complexity.

Rosmini was also able to offer a fresh form of the classical doctrine of the immortality of the soul. Immortality has its basis in the fundamental sentiment of the idea of being; through the corporeal sentiment, the body shares immortality.

## THE PERSON

Important both in itself and for its function in his social and political thought is Rosmini’s central doctrine of the person. A subject has two aspects, nature and person. A subject’s nature is the complex and sum of the activities of which the subject is agent. The perfection of the subject in the order of nature is the perfection, in number and in quality, of these activities. “Person” designates the directive unity of these activities and hence is associated in a special way with the will. The will is fundamental because it directs and organizes the activities of a person’s nature, and in so doing it exhibits the basic deontic character of the person, its orientation toward a norm, toward the ought. The person emerges as the unique and incommunicable unity of the activities of the nature through a unique and unrepeatable activity of the will. It is not merely an operational or structural unity but a deontic one, basically oriented toward the world of values and

norms and hence constitutively moral. The central effort of life is the realization of the developed or explicit person, which is achieved through the exercise of moral decision within the context of nature and its diverse activities. This effort is the basis of Rosmini’s distinction between *vita diretta* and the *vita riflessa*, which is central to his moral philosophy. The central effort of the moral life is the practice of the *vita riflessa*, the examined life in a creative sense.

The elaboration of the notion of the person gives structure to Rosmini’s moral philosophy; his philosophy of right, law, and state; and his theory of education.

## ETHICS

Personalism enabled Rosmini to overcome the formalism of Kantian ethics. The idea of being is the criterion of the good as well as of truth. In the intimate unity of the person, the speculative act of intellectual perception immediately translates itself into a practical judgment that becomes the legislative principle of action. The truth of being that intellectual perception presents inevitably involves the assenting activity of the will. The will seeks the being of all things in the idea of being, revealed in the deontic order as the good. Rosmini, on Kant’s model, tried to distill this insight into a rule: Recognize in action or practice what you have recognized speculatively. The essence of the moral life resides in this act of recognition, reflected prismatically in all the concrete situations in which the agent discovers himself. The obligatoriness of the rule springs from the fact that a hiatus between the speculative and the practical orders, between universal recognition and individual action, is intolerable. The psychological expression of this intolerance is remorse, the characteristic state of a person who deviates from this imperative. The true form of Rosmini’s moral philosophy is embodied in another imperative: Be faithful to being; specifically, to the being that is revealed in the idea of being and which is the ground of all.

Fidelity to being was immediately translated by Rosmini into a rule of justice. The idea of being contains all the grades of being. The realm of being thus constitutes at the same time a hierarchy of values. Fidelity to being demands that the rule of justice, “Give to each its due,” be interpreted in terms of this hierarchy. How is this hierarchy of values to be apprehended? Rosmini’s reply is that it is to be apprehended through spontaneous recourse to the intellectual light, the constitutive presence in the subject of the idea of being.

## POLITICAL PHILOSOPHY

The concepts of person and justice provide the bases of Rosmini's political philosophy. Abstractly, right is the property of being, for being demands to be recognized and in doing so establishes the moral and the juridical orders. Concretely, right has its locus in the person, because of the person's ontological status as subject. In the person, right becomes a capacity to act eudaemonically, a capacity that is protected by the moral law; the same law imposes on others the obligation to recognize this capacity. Rosmini sought to bring right under the moral law in order to oppose those who would make it rest on force; he made it an endowment of the person to oppose those who would assign its origin to any other source, such as organized society in any of its forms. He distinguished innate natural rights, derived connatural rights, and acquired rights. Property, by means of which the person acquires physicomoral dominion over objects, is the chief acquired right.

While property defines the relation of the person to objects, the social bond relates him to other persons. The basis of the bond of sociality among a plurality of persons is their participation in a common intelligent principle, ideal being. Rosmini placed the forms of social life on a continuum between the terms of the most rudimentary and inclusive—membership in the human race—and the most intimate and exclusive—the conjugal relationship. Civil society falls midway on this continuum. Civil society has only a functional and not a substantive character: It does not originate rights but simply regulates the mode of their enjoyment and exercise. This provides Rosmini with his definition of the state and of government: The state is a regulatory principle of the modality of human rights. A just state achieves a balance between the common good (the good of the members distributively considered) and the public good (that of the social body considered as an organism). Abstractly, the common good is to be preferred to the public good, so as to preclude justification of acts of the state by recourse to the doctrine of "reason of state"; concretely, this preferential status is less determinate.

## BEING AND GOD

In two extensive works, the *Teosofia* (posthumously published, 1859–1874) and the *Teodicea* (1845), Rosmini drew the widest possible conclusions from his personalistic premises. The theme of the *Teosofia* is the unity of being as prior to any of its modes (the absolute metaformality of being). Being, in this sense, is the basis of all the actual and determinate forms of being and contains

within itself all of the principles of that determination *in abstracto* or *virtualiter*. It is not, however, the creative principle by which those forms are reduced to actuality. The need for a creative principle opens the way for the argument of the *Teodicea*, that God necessarily exists. God is the creative principle by which the virtuality of the order of primal being is realized in the actual and concrete modes of existence.

## EDUCATIONAL THEORY

The culmination of Rosmini's thought is considered by many to be his pedagogical theory. The guiding principle of this theory is a summary of his entire philosophy: respect for the human person as the vehicle of divine light and ideal being. Rosmini stressed the unity of educational process and also methodology. The person is the principle of integrity; education is the process of the realization of the person in this sense. The principle of this integration is religion, which gives unity of purpose, unity of doctrine, and unity of powers to the educational process. The supreme law of method, the principle of gradation, ensures the conformity of the process of education to that of life. The process of growth and integration according to this law is from the universal to the particular. The object of the entire process is the free and realized person who fulfills himself in free association with other persons in all social forms and whose freedom rests ultimately upon his foundation in ideal being.

*See also* Augustine, St.; Augustinianism; Being; Empiricism; Ideas; Kant, Immanuel; Personalism; Philosophy of Education, History of; Rationalism; Realism; Thomism.

## Bibliography

### WORKS BY ROSMINI

*Opere*. Edited by Enrico Castelli, 28 vols. Rome, 1934–. One hundred volumes planned.

*Saggio sull'unità dell'educazione*. Milan, 1826.

*Nuovo saggio sull'origine delle idee*. Rome, 1830.

*Teodicea*. Milan, 1845. Translated as *Theodicy*. 3 vols. New York, 1912.

*Psicologia*. 2 vols. Novara, 1850. Translated as *Psychology*, 3 vols. 1884–1888.

*Del principio supremo della methodica e di alcune sue applicazioni in servizio della umana educazione*. Turin, 1857.

*Teosofia*. 5 vols. Turin, 1859–1874.

### WORKS ON ROSMINI

#### *Bibliographies*

*Antologia rosminiana*, Vol. 1 (1955).

Caviglione, Carlo. *Bibliografia delle opere di Antonio Rosmini*. Turin: Paravia, 1925.

Palhories, Fortunat. *La philosophie de Rosmini*. Paris, 1908.

### Biographies

Bertoletti, C. G. *Vita di Antonio Rosmini*. Turin, 1957.

Leatham, Claude. *Rosmini*. London and New York: Longmans, Green, 1957.

Pagani, G. B. *Vita di Antonio Rosmini*, 2 vols. Turin, 1897. Revised by G. Rossi and G. Bozzetti. Rovereto, 1957. Translated as *The Life of Antonio Rosmini-Serbati*. London: Routledge & Sons, 1907.

### Philosophical Studies

Bruno, G. F. "Rosmini's Contributions to Ethical Philosophy." In *Archives of Philosophy*. New York: Columbia University Press, 1916.

Bulferetti, L. *Antonio Rosmini nella Restaurazione*. Florence: F. Le Monnier, 1942.

Capone Braga, Gaetano. *Saggio su Rosmini: Il mondo delle idee*, 2nd ed. Florence, 1924.

Caviglione, Carlo. *Il Rosmini vero*. Voghera, Italy, 1912.

Chiavacci, C. *Il valore morale nel Rosmini*. Florence, 1921.

Franchi, A. *Saggio sul sistema ontologico di Antonio Rosmini*. Milan, 1953.

Galli, Gallo. *Kant e Rosmini*. Città di Castello: S. Lapi, 1914.

Gentile, Giovanni. *Rosmini e Gioberti*, 2nd ed. Florence: Sansoni, 1955.

Gonella, G. *La filosofia del diritto secondo Rosmini*. Rome: Studium, 1934.

Honan, U. *Agostino, Tommaso e Rosmini*. Milan: Sodalitas, 1955.

Morando, Dante. *La pedagogia di Antonio Rosmini*. Brescia: La scuola, 1948.

Morando, Giuseppe. *Esame critico delle XL proposizioni rosminiani condannate*. Milan: Cogliati, 1905.

Nicola, G. B. *Saggi di scienza politico di Antonio Rosmini*. Turin, 1933.

Pignoloni, E. *Il reale nei problemi della Teosofia di Antonio Rosmini*. Milan: Sodalitas, 1955.

Prini, Pietro. *Introduzione alla metafisica di Antonio Rosmini*. Milan: Sodalitas, 1953.

Riva, C. *Il problema dell'anima intelletiva secondo Antonio Rosmini*. Domodossola: Sodalitas, 1956.

Rovea, Giuseppe. *Filosofia e religione in Antonio Rosmini*. Milan: Sodalitas, 1951.

Sciacca, M. F. *La filosofia morale di Antonio Rosmini*. Rome: Bocca, 1938.

### Collections of Articles

*Atti del congresso internazionale di filosofia "Antonio Rosmini,"* 2 vols. Florence, 1957.

*Crisis* (1955), no. 6.

*Giornale di metafisica* (1955), nos. 4 and 5.

*Humanitas* (1955), nos. 9 and 10.

*Rivista della filosofia neoscolastica* (1955), nos. 4 and 5.

*Rivista rosminiana* (1955), nos. 3 and 4. Contains an extensive bibliography.

*Teoria* (1955), nos. 3 and 4.

**A. Robert Caponigri (1967)**

## ROSS, WILLIAM DAVID

(1877–1971)

William David Ross was a British Aristotelian scholar and moral philosopher. Sir David Ross was born in Scotland and was educated at the Royal High School in Edinburgh, Edinburgh University, and Balliol College, Oxford, where he took firsts in classical moderations and "greats." He was a fellow of Merton College from 1900 to 1902, when he was elected a fellow and tutor of Oriel. He was provost of Oriel from 1929 until his resignation in 1947.

Ross was prominent in academic and public life. He was vice-chancellor of Oxford University (1941–1944), pro-vice-chancellor (1944–1947), president of the Classical Association (1932), and president of the British Academy (1936–1940). He was chairman of Council of the Royal Institute of Philosophy continuously since 1940. In 1947 he served as president of the Union Académique Internationale.

Ross was awarded the Order of the British Empire for his work in the ministry of munitions and as a major on the special list during World War I. He was knighted in 1938. During World War II he was a member of the appellate tribunal for conscientious objectors and after the war was honored by the governments of Norway and Poland. Among his many public services were the chairmanships of three government departmental committees (1936–1937) and of the civil service arbitration tribunal (1942–1952). From 1947 to 1949 he was chairman of the important Royal Commission on the Press.

The qualities that made Ross successful in public life are those to which he owes his distinction as a philosopher. He was not only an Aristotelian scholar, but he also had an Aristotelian frame of mind—moderate, critical, balanced, thorough, and, above all, judicious. He valued and possessed what Aristotle called "practical wisdom" no less than speculative ability.

Ross edited the Oxford translations of Aristotle, published between 1908 and 1931. He translated the *Metaphysics* and the *Ethics* himself, and he published definitive editions of a number of Aristotle's works. His *Aristotle* (London, 1923) is mainly expository, each chapter being concerned with a major aspect of Aristotle's work; this is still the best all-round exposition in English.

Ross was the leading opponent of the view of John Burnet and A. E. Taylor that the Socrates of Plato's dialogues is never a mouthpiece for Plato's own doctrines. In *Plato's Theory of Ideas* (Oxford, 1951), Ross rejected their contention that the theory of Ideas was originally the work of Socrates and not of Plato. This book traces the

development of the theory of Ideas through Plato's thought. It includes a detailed discussion of Plato's cryptic doctrine of "ideal numbers," using Aristotle's account in the *Metaphysics* as a guide to the interpretation of the doctrine.

Ross's main contribution to philosophy, as distinct from philosophical scholarship, is in the field of ethics. In *The Right and the Good* (Oxford, 1930), he argued the case for intuitionism with a lucidity and thoroughness that made the book a classic. For some ten years it was the center of ethical controversy. In his *Foundations of Ethics* (Oxford, 1939) Ross restated his case and replied to his critics.

Ross's approach to ethics is Aristotelian. "The moral convictions of thoughtful and well-educated people are the data of ethics, just as sense-perceptions are the data of a natural science" (*The Right and the Good*, p. 41). He appeals to what we mean by rightness and goodness and assumes that this guarantees the existence of what is meant and is a sure guide to its nature.

The germ of Ross's position is to be found in an article by H. A. Prichard, "Does Moral Philosophy Rest on a Mistake?" (*Mind* 21 [1912]: 21–152; reprinted in *Moral Obligation*, Oxford, 1949, pp. 1–17). Prichard was a pupil of John Cook Wilson, who also influenced Ross directly, an influence that appears in Ross's opposition to reductionism and in his view that knowledge and opinion are distinct in kind. The other main debt acknowledged by Ross is to G. E. Moore, whose arguments against ethical subjectivism he endorses, although he rejects Moore's "ideal utilitarianism."

Right and good are for Ross distinct, indefinable, and irreducible objective qualities. Rightness belongs to acts, independently of motives; moral goodness belongs to motives. Ross uses "act" for what is done and "action" for the doing of it. Thus, the doing of a right act may be a morally bad action—that is, a right act can be done from a morally bad motive; the inverse also holds. Nor can it ever be morally obligatory to act from a good motive. There are four kinds of good things—virtue, knowledge, pleasure, and the allocation of pleasure and pain according to desert. No amount of pleasure equals the smallest amount of virtue. In *Foundations of Ethics* Ross argued that virtue and pleasure are not good in the same sense—virtue is "admirable," pleasure only "a worthy object of satisfaction." What alone is common to the two senses is that they express a favorable attitude.

Ross's two main targets are ethical subjectivism and "ideal utilitarianism," which "ignores, or at least does not

do full justice to, the highly personal character of duty" (*The Right and the Good*, p. 22). Specific duties are of three kinds—reparation, gratitude, and keeping faith. The "plain man" (to whom Ross, as a good Aristotelian, frequently appeals), in deciding what he ought to do, thinks as often of the past (a promise made, a debt incurred) as of future consequences. Ross does, however, admit among duties the utilitarian general duty of beneficence when it does not conflict with a specific duty. And "even when we are under a special obligation the tendency of acts to promote general good is one of the main factors in determining whether they are right" (p. 3a).

Conflict of duties is one of the main problems facing an intuitionist, who cannot accept the utilitarian's "Do what will produce the most good." Ross says: "Do whichever act is more of a duty." To make sense of "more of a duty," he draws a distinction between prima-facie and actual duties and holds that conflict can only arise between prima-facie duties. An act is a prima-facie or "conditional" duty by virtue of being of a certain kind (for instance, the repaying of a debt) and would be an actual duty if it were not also of some other morally important kind or did not conflict with another more important prima-facie duty. Thus, if I have promised to lend money to a friend in need, I have a prima-facie duty to hand over the money. But suppose that before I have done so, I find that I need it for the legal defense of my son, charged with a crime of which I believe him innocent. I recognize a conflicting prima-facie duty to help him. Ross maintains that (a) one, and only one, of these two prima-facie duties is my actual duty; (b) I know each of them to be a prima-facie duty—this is self-evident; (c) I can have only an opinion about which is "more of a duty" and therefore my actual duty.

**See also** Aristotelianism; Aristotle; Ethical Subjectivism; Ethics, History of; Intuitionism; Intuitionism and Intuitionistic Logic; Moore, George Edward; Plato; Socrates; Taylor, Alfred Edward; Utilitarianism.

## Bibliography

### OTHER WORKS BY ROSS

*Kant's Ethical Theory: A Commentary on the Grundlegung zur Metaphysik der Sitten*. Oxford, Clarendon Press, 1965.  
*The Right and the Good*, edited by Philip Stratton-Lake. Oxford: Clarendon, New York: Oxford University Press, 2002.

Some works on Ross are J. H. Muirhead, *Rule and End in Morals* (London: Oxford University Press, H. Milford, 1932); H. J. McCloskey, "Ross and the Concept of Prima Facie Duty," in *Australasian Journal of Philosophy* 41 (1964): 336–345; G. E. Hughes, "Motive and Duty," in *Mind* 53

(1944): 314–331, W. A. Pickard-Cambridge, “Two Problems about Duty,” in *Mind* 41 (1932): 72–96, 145–172, 311–340; P. F. Strawson, “Ethical Intuitionism,” in *Philosophy* 24 (1949): 23–33.

A. K. Stout (1967)

*Bibliography updated by Michael J. Farmer (2005)*

## ROUGIER, LOUIS

(1889–1982)

Louis Rougier, the French philosopher, was a pupil of Edmond Goblot. Rougier taught philosophy at the universities of Besançon and Caen. In 1935 he organized and presided over the Paris International Congress of Scientific Philosophy, where the leading spokesmen for logical empiricism, at the time little known in France, presented their views in a body.

From the start, Rougier’s thought had been marked by the contemporary upheavals in the sciences of physics, mathematics, and logic. To these developments he devoted several of his early books, including *La philosophie géométrique d’Henri Poincaré* (Paris, 1920), *La structure des théories déductives* (Paris, 1921), *La matière et l’énergie selon la théorie de la relativité et la théorie des quanta* (Paris, 1921), and *En Marge de Curie, de Carnot et d’Einstein* (Paris, 1922).

In his view, the upsets in the sciences reinforced the closely pressed critique which he had directed in his doctoral thesis, *Les paralogismes du rationalisme* (Paris, 1920), against the theory academic philosophers call “rationalism.” This is an a priori rationalism, quite different from scientific and experimental rationalism. It asserts the existence of a universal, immutable reason and of eternal, necessary truths, with all the theological, ontological, and epistemological implications that such a thesis requires. According to Rougier, the body of notions and principles that constitute “reason” in the classic sense is simply the characteristic of a certain mental structure, the ontological or metaphysical temperament, which is also the subject of his detailed study *La scolastique et le thomisme* (Paris, 1925). Besides the temperament dominated by “reason,” history discloses other temperaments—animistic, symbolic, scientific—having command of other types of explanation. The human mind possesses an infinite plasticity; it is able to take delight in quite varied forms of intelligibility, without any internal necessity having compelled it to evolve in just the direction that it has. If the laws of logic are necessary truths, it is only because they are tautologies in the sense

of Ludwig Wittgenstein; that is, they are devoid of any information about the universe and hence stripped of any ontological import. Even this logical necessity, as is shown by the existence of a plurality of logics, is relative to a given system of axioms and rules.

This rejection of all a priori synthesis, this radical separation between logico-mathematical statements and empirical statements, and the condemnation it entails of all metaphysics as victim of the imperfections of our natural languages (*La métaphysique et le langage*, Paris, 1960), closely ally Rougier’s philosophy to that of logical empiricism. His long *Traité de la connaissance* (Paris, 1955) offers analyses illustrated with abundant examples from the past and contemporary history of the sciences; in style and ideas it is probably closer than any other French book to the majority of central European and American works on epistemology. Nevertheless, certain features testify to his originality in comparison with the logical empiricism of the Vienna circle. Rougier rejects the physicalist reduction and upholds a plurality of languages. Nor does he agree that all basically unsolvable problems must by their nature alone be regarded as devoid of meaning; besides, meaninglessness is a notion relative to the language chosen. Further, several of the ideas he developed in works other than the *Traité*, for example his thesis of the diversity of mental structures and the plasticity of the intellect, do not strictly belong to the common stock of the school of logical empiricism, but have been added to it.

Although epistemology and the critique of knowledge are at the center of Rougier’s philosophy, he wrote in two other fields. One is the history of scientific, philosophical, and religious ideas, to which he devoted *Celse ou le conflit de la civilisation antique et du christianisme primitif* (Paris, 1926) and *La religion astrale des Pythagoriciens* (Paris, 1959). The other is contemporary political problems; he dealt critically with the democratic and egalitarian ideology of the “men of 1789” and their successors in such works as *Les mystiques politiques et leurs incidences internationales* (Paris, 1935), *Les mystiques économiques* (Paris, 1949), and *L’erreur de la démocratie française* (Paris, 1963).

Rougier systematically omitted these two aspects of his thought from the account he himself gave of his “philosophical itinerary” (*La revue libérale*, no. 33, [1961]: 6–79), an account which can well serve as an overall study of his theory of knowledge.

**See also** A Priori and A Posteriori; Epistemology; Logical Positivism; Rationalism.

## Bibliography

There are no studies of Rougier, but the reader may wish to consult the brief account by A. D. Dandieu in *Anthologie des philosophes français contemporains* (Paris: Éditions du Sagittaire, 1931) followed by a detailed bibliography and two extracts from Rougier's writings.

### WORKS BY ROUGIER

*Celse discours vrai contre les Chrétiens*. Paris: J.-J. Pauvert, 1965.  
*Histoire d'une faillite philosophique: La Scolastique*. Paris: J.-J. Pauvert, 1966.

*Le culte des images et la primitive Église*. Paris: (7e), Cercle Ernest-Renan, 3, rue Récamier, 1970.

*The Genius of the West*. Los Angeles, CA: Nash, 1971.

*La genèse des dogmes chrétiens*. Paris: A. Michel 1972.

*La métaphysique et le langage; Essai*. Paris: Denoël 1973.

*Le conflit du christianisme primitif et de la civilisation antique*. Paris: Grece, 1974.

*Celse contre les chrétiens: La réaction païenne sous l'empire romain*. Paris: Copernic, 1977.

*De Gaulle contre De Gaulle*. Paris: Éditions du Triolet 1978.

*Du paradis à l'utopie*. Paris: Copernic, 1979.

*Astronomie et religion en Occident*. Paris: Presses universitaires de France, 1980.

*La mystique démocratique: Ses origines, ses illusions*. Paris: Éditions Albatros, 1983.

*La religion astrale des pythagoriciens*. Monaco: Éditions du Rocher, 1984.

Poincaré, Henri. *Scientific Opportunism: An Anthology = L'opportunisme scientifique*. Compiled by Louis Auguste Paul Rougier; edited by Laurent Rollet. Boston: Birkhäuser, 2002.

**Robert Blanché (1967)**

Translated by Albert E. Blumberg

Bibliography updated by Michael J. Farmer (2005)

## ROUSSEAU, JEAN-JACQUES (1712–1778)

Jean-Jacques Rousseau, the philosopher, essayist, and novelist, was born at Geneva. His mother having died a few days after his birth, he was brought up by an aunt and an erratic father who taught him to read through the medium of sentimental novels and Plutarch's *Lives*. He had little formal education. After staying for about two years with a country minister at Bossey, he returned to Geneva and lived with an uncle. He was then apprenticed in turn to a notary and an engraver, the latter of whom treated him so brutally that in 1728 he left Geneva to seek his fortune elsewhere.

Rousseau was protected and befriended by Mme. de Warens, a convert to Roman Catholicism, who had left her native canton of Vaud to live at Annecy in Savoy, with financial support from the king of Sardinia and the ecclesiastical authorities. Rousseau's subsequent attachment to

her was a decisive factor in his conversion to Roman Catholicism as well as in his emotional development. He made a formal abjuration of Protestantism at the hospice for catechumens at Turin. He then served for a time as a lackey, finally returning to Mme. de Warens in 1729. Thereafter, he led an unsettled life, restless travel alternating with a more stable existence at Chambéry, where Mme. de Warens had established herself. Intellectually, the most important event of this phase of his life was a protracted spell of enthusiastic study under his own direction. A brief experience as a private tutor at Lyons in 1740 helped to create a lifelong interest in education and at the same time convinced Rousseau that he had no aptitude for this profession. As he had acquired some musical competence at Annecy, he set out hopefully for Paris in 1742 with a new system of musical notation. Although this did not bring him the success he hoped for, he was introduced to a number of influential people, including the wealthy Mme. Dupin and her stepson M. de Francueil.

In 1743, Rousseau was appointed secretary to the French ambassador at Venice, M. de Montaigu, but he lost this post the following year because of a quarrel with him. On his return to Paris, Rousseau increased his difficulties by an irregular union with an ignorant servant girl, Thérèse Le Vasseur, in 1745; by her he probably had five illegitimate children, who were all sent to a foundlings' home. He also met Denis Diderot, Jean Le Rond d'Alembert, and other *philosophes* and was invited to contribute musical articles to the *Encyclopédie*.

Rousseau's literary career began in 1750 with the publication of the *Discours sur les sciences et les arts*, which had previously won a prize at the Academy of Dijon. However, his first real success came with the performance of his opera *Le devin du village* before Louis XV at Fontainebleau, but his refusal to allow himself to be presented to the king lost him any chance of securing a royal pension. A journey to Geneva in 1754 led to a reconciliation with the republic and a formal return to Protestantism.

After the publication in 1755 of his *Discours sur l'origine de l'inégalité*, Rousseau felt increasingly unhappy in Paris, and in 1756 he installed himself in a small country house, called the Hermitage, which belonged to a rich friend, Mme. d'Épinay. There followed a comparatively short but intense period of literary activity that saw the publication of the *Lettre à d'Alembert sur les spectacles* (1758), *Julie, ou la Nouvelle Héloïse* (1761), *Émile* (1762), and the *Contrat social* (1762).

During this time Rousseau's relations with the Encyclopedists became increasingly strained, with intellectual differences, especially on the subject of religion, being aggravated by personal quarrels with former friends such as Diderot and the Baron von Grimm. In 1762 the condemnation of *Émile* by the Paris Parlement forced him to flee from France and settle in Neuchâtel under the protection of the king of Prussia. In the *Lettre à M. de Beaumont* (1763) Rousseau vigorously defended the "Profession de foi du vicaire savoyard," which had been included in the fourth book of *Émile*, against its condemnation by the archbishop of Paris; this was followed in 1764 by another polemical work, the *Lettres écrites de la montagne*, provoked by increasing opposition from the Genevan authorities to his political and religious views. Alarmed by local hostility, Rousseau decided to leave the Neuchâtel region in 1765, and he accepted an invitation from the philosopher David Hume to make his home in England. His arrival in that country in 1766 and his subsequent residence in Derbyshire were disturbed by the appearance of abnormal emotional and mental reactions, culminating in the irrational conviction that Hume's invitation had been a mere pretext for Rousseau's defamation. After quarreling violently with Hume (who riposted by publishing an account of the affair), Rousseau fled panic-stricken to France in 1767. For the next few years he moved from place to place, oppressed by the thought of universal persecution. He eventually settled in Paris in 1770 and died suddenly on July 2, 1778, less than two months after he had gone to live on the estate of the marquis de Girardin at Ermenonville.

The chief literary activity of Rousseau's last years was the composition of a remarkable series of personal works, the *Confessions*, on which he had worked intensively during his stay in England; the dialogues known as *Rousseau juge de Jean-Jacques*, a curiously pathological document illuminated by some pages of remarkable brilliance and insight; and the beautiful but unfinished *Rêveries du promeneur solitaire*. These writings are remarkable for their lyrical power and sustained efforts at self-analysis.

## THOUGHT

From the very first Rousseau's work betrayed the strongly personal emphasis of a writer who felt that he did not truly belong to his immediate environment. Being of Genevan origin, largely self-taught, and endowed with a particularly sensitive temperament, he could never bring himself fully to accept the social and moral implications of French culture, even though he never ceased to admire French taste. In 1749, as he was on his way to Vincennes

to visit his imprisoned friend Diderot, he saw in a copy of the *Mercure de France* the subject of the prize essay set by the Academy of Dijon: whether the restoration of the arts and sciences has contributed to the purification of manners. In the *Confessions* he writes that at that moment he experienced a sudden "illumination" and "inspiration," the dazzling vision of a "new universe," which impelled him to answer the academicians' question with an emphatic "No!" Although this viewpoint was already familiar to a certain type of traditional Christian moralist, Rousseau struck a new personal note remarkable for its deeply felt sincerity; he always refused to consider himself as a professional man of letters and stressed his role as an independent writer with a message for humanity.

**NATURE AND SOCIETY.** Rousseau's early works (the two discourses and the *Lettre à d'Alembert*) developed the fundamental antithesis that he deemed to exist between contemporary society and the nature of man. European civilization was indicted for having sacrificed the moral demands of human nature to the superficial allure of a purely intellectual culture and thus for having replaced natural by artificial needs. The artificial uniformity of behavior that society imposes on people causes them to ignore "the duties of man and the needs of nature," so that appearance and reality are constantly at variance in modern social life, as for example in the case of an excessive regard for politeness and convention concealing the most ruthless and calculating egoism. Likewise, insisted Rousseau, the sciences and the arts, in spite of their brilliance, are not the genuine expression of fundamental human needs but the result of pride and vanity. The rapid growth of luxury and idleness serves merely to increase the corruption of the contemporary situation. Consequently, as culture appears to attain an ever increasing splendor, genuine human relationships become steadily weaker. Man is alienated from his original nature and prevented from being his real self; a perpetual prey to inner contradictions, he vainly grasps at objects outside himself as he neglects the true lessons of nature in order to pursue the illusions of opinion.

To "society" Rousseau opposed "virtue"—a constant theme of his early works. Virtue confers stability and unity upon human existence because it subordinates idle speculation to the active needs of the moral life. Unlike mere reflection, it induces "strength and vigor of soul," allowing full expression to man's genius and conferring on his existence a solidity and permanence that are quite unlike the ephemeral brilliance of contemporary culture. Whereas society forces man to assume the mask of



hypocrisy and deceit as a means of satisfying his selfish interests, virtue, “the sublime science of simple souls,” gives him an authentic openness and innocence that allow him to reveal himself to others as he truly is.

A particularly serious feature of modern society is the prevalence of an unnatural inequality based on power and wealth. In the *Discours sur l'origine de l'inégalité* Rousseau examines this phenomenon in the light of man's evolution from the primitive state to his present existence as a political being and concludes that modern conditions represent a fall from happiness into misery. In spite of its historical form, this discourse, as the author himself admits, is a purely hypothetical and imaginative reconstruction that deliberately ignores facts, whether historical or theological, in order to concentrate on the nature of man as it is revealed to Rousseau's intuitive perception. If the state of nature can never be known as a historical fact, it at least serves as a useful concept that enables him to distinguish man's original qualities from fortuitous historical accretions.

Limited and instinctive though the life of primitive man may have been, it was at least a happy one inasmuch as the savage knew how to live in accordance with his own innate needs. Leading an isolated existence in the forests, satisfying his basic appetite for food and sex without difficulty, untouched by modern man's anxiety before illness and death, he was largely self-sufficient; the primordial urge toward self-preservation was effectively counterbalanced by an innate feeling of natural pity that prevented him from inflicting needless pain upon his fellow men. Man was from the outset endowed with free will and perfectibility, but these became active only when the first rudimentary social communities, based mainly on the family, were established, a period that Rousseau treats as the golden age of humanity since it lay halfway between the brutishness of primitive existence and the corruption of political societies. The discovery of agriculture and metallurgy and the distinction between “mine” and “thine” meant that people had to work together, and this inevitably led to the establishment of property. Men then became divided into rich and poor and, later, into powerful and weak, so that the inequality of the social system was at last made permanent through the institution of laws and political organization. In Rousseau's opinion the historical process will culminate in the triumph of despotism, which makes all men once again equal because all have become slaves of one master.

Whereas the early discourses dealt mainly with general principles, the publication of d'Alembert's article “Geneva” in the seventh volume of the *Encyclopédie* in

1757, with its suggestion that the Genevans would benefit from the establishment of a theater, led Rousseau to deal with a specific aspect of his criticism of society. In his various replies to early critics he had already insisted that man, having once left the primitive state, could never return to it; he also maintained that it was the large states, especially the monarchies of Europe, which had traveled furthest on the road to perdition. However, small republics like Geneva, though no longer close to nature, still retained a relative simplicity and innocence and could be protected against further corruption. To introduce the theater into Geneva was, in Rousseau's eyes, to bring an evil product of society into a comparatively unspoiled community. The *Lettre à d'Alembert* also provided him with an opportunity of examining not only the general characteristics of the theater but also the whole question of what amusements are best suited to man's true nature.

Starting from the assumption that all valid entertainment must “derive from man's work, relationships and needs,” Rousseau insists that it must be an integral part of man's daily life, different from his work and yet inspired by the same spirit. The theater, however, is primarily an artificial entertainment, the product of idleness and vanity and the fomenter of dangerous passions and emotions; it is always subservient to the impulses that create it and remains incapable of directing people toward moral activity. The theater is typical of a large city like Paris, with its reversal of natural values. Whereas for Rousseau woman is naturally modest and self-effacing, the theater makes her a shameless figure who transforms love into a public spectacle; the very existence of actresses also sets the example of a completely unfeminine way of life that is characteristic of a society in which women set the tone and rule the *salons*, reducing men to a condition of abject and effeminate dependence. By contrast, Rousseau extols the simplicity of the Montagnons, the simple, industrious mountain dwellers whom he remembered from his youth and recalls with heartfelt enthusiasm. Unlike modern men such people relied upon their own creative resources for their work and entertainment. The Genevans, too, through their “societies” and “circles,” wisely allowed men and women to indulge in their own separate pastimes. The *Lettre* ends with a remarkable evocation of the kind of national entertainment that, in Rousseau's opinion, would be suitable for a small homogeneous community like Geneva. The Genevans should actively participate in a joyous public entertainment that takes place “beneath the sky” and in the presence of their fellow citizens; in this way the whole community would be inspired by feelings that are both social and human.

Perhaps one of the gravest general aspects of society's harmful influence on the nature of man is its constant tendency to transform *amour de soi* (self-love) into *amour-propre* (pride). Although this antithesis was not peculiar to Rousseau, who had already noted its existence in Vauvenargues, it does, occupy a particularly significant position in his social criticism. *Amour de soi* is always good and, in its purest state, quite spontaneous because it expresses the real essence of human existence. It is an absolute feeling or passion that serves as the source of all genuinely natural impulses and emotions; already revealing itself at the instinctive level as the desire for self-preservation, it assumes a much nobler expression as soon as it is combined with reason. Being in complete uniformity with the principle of order, it will affect all the main aspects of human existence as it brings the individual into contact with his own inner self, his physical environment, and his fellow men. Unfortunately modern society has changed this natural *amour de soi*, which makes a man what he truly is, into *amour-propre*, an artificial reaction originating in an anxious reflection that induces a man to be forever comparing himself with others and even finding his sole pleasure in their misfortune or inferiority; through *amour-propre* he is taken outside himself into the realm of illusion and opinion and so prevented from being a complete person.

EDUCATION. Having diagnosed the malady of modern civilization, Rousseau was faced with the task of suggesting a cure, and this led him into the domain of education and politics, activities that are, or should be, rooted in man's moral nature. Rousseau was convinced that man's original nature is good, but that it has been corrupted mainly by the historical accident of society. It therefore seemed quite consistent to affirm that men are wicked but that man himself is good. To be good is to exist in accordance with the intrinsic potentialities of one's nature, and *Émile* seeks to trace the natural development of a human being brought up in the country away from the nefarious influence of contemporary social life. From this point of view the work is not just a manual of education but, as Rousseau himself points out, a philosophical treatise on the goodness of human nature. It is less concerned with laying down the practical details of a specific pedagogic method than with describing the fundamental principles that underlie the whole of man's development from infancy to maturity. Rousseau's ultimate object is to teach the art of living, for man's first duty, he says, is to be human.

The educator must realize that "vice and error, alien to man's constitution, are introduced into it from out-

side"; his first task will be to keep away harmful influences from the young child. This is why *Émile* is set in a rural environment that allows the child to grow in accordance with his own nature. Early education is therefore largely negative insofar as it is mainly concerned with removing obstacles that might hinder this development.

From the first Rousseau stresses the importance of a progressive education: Each stage of the process must be carefully adapted to the individual's developing needs and so follow "the natural progress of the human heart." In this respect Rousseau uses in his own way the genetic method of contemporary thinkers like d'Alembert, Condillac, and Comte de Buffon, who, in turn, had taken over the notion of the genealogy of ideas developed by John Locke in his famous *Essay concerning Human Understanding*. In *Émile*, however, as in the *Discours sur l'origine de l'inégalité*, Rousseau does not strive to establish an inductive law based on the empirical examination of facts but starts from a fundamental principle (man's natural goodness) that is derived initially from personal intuition, though he believes it to be subsequently verifiable by observation and psychological analysis. *Émile* therefore involves certain metaphysical elements, but these are referred back to the concrete aspects of human nature.

Rousseau maintains that a truly progressive education will recognize that the child has his own special needs as a being who exists in his own right. "Nature wants children to be children before being men.... Childhood has its own ways of seeing, thinking and feeling." Since the child's needs are largely physical, negative education "tends to perfect the organs, instruments of our knowledge." Incapable of dealing with abstractions, the child must be educated through contact with things. To him dependence on things will be natural and inevitable; acknowledging only the "heavy yoke of necessity," he will escape the tyranny of any human will. Unlike the despotic power of men necessity is quite compatible with properly controlled freedom since it lets the human being exercise his powers within the limits prescribed for him by nature. "The truly free man" wishes to do no more than this. Well-regulated freedom thus provides the only valid basis and aim of sound education.

Early education, being based primarily on the senses, ignores bookish learning for direct contact with the physical world. Learning through a process of trial and error, the child experiments, as it were, through the medium of facts rather than words. (The sole book Rousseau will allow the child is Daniel Defoe's *Robinson Crusoe* and that only because it describes a man's reliance on his own

ingenuity and resourcefulness.) Freed from the tyranny of human opinion, the child identifies himself effortlessly with the requirements of his immediate existence; content to be himself and completely absorbed in his present being, he leads a kind of self-sufficient, timeless existence that knows no anxious concern for the future, none of that tormenting foresight that causes modern man to be so unhappily “outside himself.” The child is happy because he is unaware of artificial needs or of any serious disproportion between capacity and desire, power and will, and in this respect he resembles the happy savage.

Rousseau recognizes that even at the stage of greatest inner harmony, the child must be prepared for the future, for in him there is a reservoir of potential energy that he does not immediately need. The educator’s task is to hold back this energy, this “superfluous aspect of his immediate being,” until it can be effectively used. It is particularly important to avoid any precocious excitement of the imagination that may be the source of future unhappiness. These dangers will be largely averted if, after the lesson of necessity, the child learns that of utility, his developing reason being applied only to what interests and helps him. That is why his early judgment must be formed not through words or abstractions but through sensations and feelings.

A truly positive education begins only when the child becomes aware of his relationships with other people, although these early social lessons will be based on sensibility rather than reason, in particular on the innate feeling of pity, with its later concomitants of love and aversion. There are no good or bad passions, says Rousseau. All are good when they are under our control; all are bad when they control us. Through the force of our passions we are impelled beyond ourselves; through the “superabundance of our strength” we are induced to “extend our being.” With the growth of sensibility, reason, and imagination the child leaves the self-sufficiency of the primitive stage for a fuller life involving relations with the physical realm of nature and the world of human beings. The educational process must therefore be carefully timed and controlled so that the various potentialities of the human being are brought to fulfillment in an orderly and harmonious manner.

It is clear from the last book of *Émile* that man must be educated for society, though not necessarily for society in its present form. Man’s nature is not fully mature until it becomes social. However, the natural man in the state of nature and the natural man in the social state cannot be identical, for whereas the former is predominantly an instinctive, primitive creature living on the spontaneous

expression of his innate vitality, the latter is a rational, moral being aware of his obligations to other people, a man called upon to subordinate the impulse of “goodness” to the demands of “virtue.” Therefore, only in society can a genuinely human morality become possible. If by “nature” is meant the merely primordial responses of the presocial man, then it is true to say that “good institutions denature man” inasmuch as they raise him up from the absolute self-sufficiency of the isolated primitive state to the level of a moral, relative existence based on an awareness of the common good and the need to live in harmonious relationship with his fellow men. Since morality inevitably involves the problem of man’s life as a social being, it is impossible to separate morality and politics, and Rousseau states most emphatically that “those who want to treat morality and politics separately will never understand anything about either.” This is a most important aspect of his political thinking. If “nature” intended man for a moral existence, then it also intended him for social life; indeed, only through the individual’s participation in the “common unity” can full personal maturity become possible. “Nature” is still the norm, but one that has to be re-created, as it were, at a higher level, conferring on man a new rational unity that replaces the purely instinctive unity of the primitive state.

**POLITICAL THEORY.** There appears to be no valid reason for finding, as some critics have done, any fundamental contradiction between *Émile* and the *Contrat social*. Such a difficulty arises only when anachronistic attempts are made to explain Rousseau’s thought in purely individualist or collectivist terms. If at first sight *Émile* seems to be an isolated individual, this is mainly because Rousseau wanted to stress the importance of the human being’s natural development, and it in no way excludes the idea that all true education must eventually be for society.

In itself the particular form of education, like that of government, must be determined by specific historical and physical conditions, but Rousseau was less concerned with this question than with that of the fundamental principles on which all true education and all true government must be based. In this respect *Émile* and the *Contrat social* are similar since each is a theoretical, normative work. Rousseau points out in his correspondence that the *Contrat social* is a philosophical discussion of political right (the work is actually subtitled *Principes du droit politique*) rather than an examination of any existing form of government. As he says in the introduction to his work, he is taking “men as they are” and “the laws as they can be.” He seeks to reconcile “what right permits

with what interest prescribes, so that justice and utility are not divided.” In Rousseau’s eyes this is what distinguishes his approach to political problems from Baron de Montesquieu’s. Whereas Montesquieu is concerned with “the positive right of established governments,” Rousseau, as the theorist of political right, examines the philosophical basis of all legitimate government.

Although the *Contrat social* has often been described as the forerunner of totalitarianism, this interpretation is certainly not consistent with Rousseau’s conscious intention, for from the very outset his overriding preoccupation is the same as it was in *Émile*—the problem of freedom. No doubt, just as the concept of nature undergoes a radical transformation when it is applied to society, so the natural freedom enjoyed by man in the state of nature differs in important respects from the civic freedom of the social state; both, however, are natural to man at different stages of his development. Man living in society faces a problem that does not affect primitive man—namely, the possible tyranny of his fellow men. Now, a true and just society can never be based on sheer force, for right can never be equated with might. Rousseau vigorously repudiates traditional views that seek to justify the right of conquerors to subject the vanquished to permanent enslavement; no society founded on such a principle can ever be legitimate. Man’s participation in society must be consistent with his existence as a free and rational being. Society is therefore unthinkable without a freedom that expresses man’s most fundamental attribute. “To give up freedom is to give up one’s human quality: to remove freedom from one’s will is to remove all morality from one’s actions.” Moreover, it is with the emergence of society that man comes into possession of his freedom and thus attains the status of a moral being. The institution of any genuine political society must be the result of a social pact, or free association of intelligent human beings who deliberately choose to form the type of society to which they will owe allegiance; this is the only valid basis for a community that wishes to live in accordance with the requirements of human freedom.

However, there still remains the problem of finding a form of association that will continue to respect the freedom that brought it into being. Although man is naturally good, he is constantly threatened by forces that not only alienate him from himself but also transform him into a tyrant or a slave. From this point of view the political problem is not dissimilar from the pedagogic one. How is man to be protected from the tyranny of the human will? Just as the child has to be liberated from dependence upon human caprice in order to confront

necessity, so the individual is to be preserved from tyranny by “an excessive dependence” of all citizens on a new kind of necessity, on something that is greater than the citizen himself and yet in one sense a part of his life. Rousseau seeks a form of association in which “each one uniting with all obeys, however, only himself and remains as free as before.” In other words, “each one giving himself to all gives himself to nobody.” The possibility of inequality and injustice will be avoided through the “total alienation of each associate, with all his rights, to the community”; if such alienation were less than total, it would expose the individual to domination by others. As it is, the citizen does not obey some sectional interest but the general will, which is a “real force, superior to the action of any particular will.” Nor, in Rousseau’s view, need this arouse any apprehension, for unlike the individual will which concerns itself with specific and perhaps selfish interests, the general will is always directed toward the general good.

Moreover, total alienation involves equality in another way; the general will is not simply an external authority that the citizen obeys in spite of himself but the objective embodiment of his own moral nature. In accepting the authority of the general will, the citizen not only belongs to a collective, moral body but also achieves true freedom by obeying a law that he has prescribed for himself. Through the law he escapes from the bondage of appetite in order to follow, as an intelligent being, the dictates of reason and conscience. Submission to a will possessing an “inflexibility which no human force could ever overcome” leads to a freedom that “keeps a man exempt from vice” and to “a morality which lifts him up to virtue.” The individual is thereby invested with another kind of goodness, the genuine virtue of the man who is not an isolated being but part of a great whole. Liberated from the narrow confines of his own being, he finds fulfillment in a truly social experience of fraternity and equality with citizens who accept the same ideal.

This conception of political right is essentially democratic insofar as the source of all political authority and, therefore, of all true sovereignty must always lie with the people as a whole. Moreover, such sovereignty is both inalienable and indivisible since, as the basis of freedom itself, it is something that can never be renounced by the people or shared with others. However, Rousseau establishes an important distinction between sovereignty and government. The sovereign, or subjects (for “sovereign and subjects are simply the same people in different respects”), may delegate the executive function of the state to the prince, or government, which thus becomes

the agent, or officer, of the people; this is true whatever the form of any particular government, whether monarchy, aristocracy, or republic. If every legitimate government is democratic in essence, this does not mean that democracy, as a definite political institution through which the people themselves carry on the government by assembling as a body, is either possible or desirable in modern conditions. Any specific form of government, as Rousseau was to show very clearly in his *Projet de constitution pour la Corse* (1765) and his *Considérations sur le gouvernement de la Pologne* (probably written about 1770–1771), will depend on a variety of historical and geographical factors.

Law, as the act of the general will and the expression of sovereignty, is of vital importance, for the establishment of sound laws can determine the whole destiny of the state. As Rousseau observes, only the gods themselves would be capable of giving good laws to the human race. That is why the legislator has such an important role in the *Contrat social*; he is invested with a remarkable, almost divine quality. It is from him that the citizen “receives in some way his life and his being”; through the legislator’s actions he experiences a genuine transformation of his personal life, forsaking the “physical, independent existence he received from nature” for a moral existence as a social being. This new mode of existence is not something imposed upon him from the outside but a possibility elicited from the depths of his inner self. The legislator is in one respect an almost godlike figure, but his purpose is to serve the essential needs of human nature.

At the end of the final version of the *Contrat social* (though not in the original draft), Rousseau seems to acknowledge that an even more powerful sanction may be required to ensure complete political stability, for he proposes to introduce into the state a kind of civil religion or civic profession of faith to which every citizen, having once given his free assent, must remain obedient under pain of death. This is an aspect of Rousseau’s political thought that many commentators have found either shocking or inconsistent. However, it will already be clear that Rousseau is no liberal in the classical political sense since he does not believe in the possibility of any rigid separation of the individual and the state; the development of a full moral life is inconceivable without active participation in society, and the unity and permanence of the state depend, in turn, upon the moral integrity and undivided loyalty of its citizens. This civic profession of faith is deliberately restricted to the “few simple dogmas” that, according to Rousseau, every rational, moral being

ought readily to accept: belief in a supreme being, the future life, the happiness of the just, and the punishment of the wicked, together with a “single negative dogma, the rejection of intolerance.” Anybody repudiating these principles would presumably be, in Rousseau’s opinion, little more than a criminal who, by forfeiting his right to be considered as a responsible human being, threatens the state with anarchy and dissolution. The practical implications of this view may still sound alarming to a modern liberal, but they are not necessarily inconsistent with Rousseau’s ideas.

**RELIGION.** If the chapter “Civil Religion” seems to strike a new note in the *Contrat social*, it is certainly not incompatible with the religious emphasis of Rousseau’s thought, for religion had always played an important role in his work, as the “Profession de foi du vicaire savoyard” made clear. Nature itself must be understood in the widest sense, as the whole realm of being originally created by God, who guarantees its goodness, unity, and order. Rousseau offended traditional Christian orthodoxy with his belief that man needs no intermediary between himself and God and is able to attain salvation by his own efforts. (In spite of his great respect for the figure of Jesus and the message of the Gospels, Rousseau could not accept the notion of the Incarnation as a solution to the problem of human sin.) But Rousseau never doubted the importance of accepting God’s existence; man, he believed, is impelled toward God by the evidence of both feeling and reason, for apart from the presence of intelligence in the universe there is also the sensitive man’s deep “feeling for nature” and the inescapable conviction of a real bond uniting his immortal soul with the spiritual order that underlies the outward appearance of the physical world.

As is well known, Rousseau was the eighteenth-century writer who gave particularly eloquent expression to this aspect of the “feeling for nature.” Furthermore, apart from the testimony of reason and sensibility there is also that of the all-powerful conscience, the “divine instinct” or “voice of the soul” which forms the basis of man’s moral existence. In moments of doubt and perplexity, when all else fails man, he can always rely for guidance on the promptings of his conscience. This does not mean that reason is thereby excluded, for reason is to be condemned only when it becomes the instrument of blind passion or selfish reflection—in other words, when it fails to recognize its dependence upon other essential elements of human nature. Conscience, however, is an even more important attribute; it is a fundamental feeling that is strikingly effective when reason may be impotent.

Even so, conscience, reason, and freedom are all integral elements of man's natural endowment, potentialities that it is his right and duty to develop, for God gave him "conscience to love the good, reason to know it and freedom to choose it." It is only through the harmonious development of all man's faculties that he can come to a full understanding of his own nature and the place allotted to him by God in the universal order.

At first sight Rousseau's philosophy seems to retain many characteristics of the traditional metaphysical outlook, and several critics have stressed his great admiration for Plato and Nicolas Malebranche. In Rousseau's eyes the universe still possesses a rationality, order, and unity that reflect the wisdom and intelligence of its creator. Yet this cannot be known by reason alone, for although reason has a function in all reflection about the meaning of the world, the heart may often provide surer insights into the ultimate mystery of creation. Moreover, Rousseau's system took the form of a series of basic intuitions that he subsequently linked together into a unified whole. His thought, therefore, is imbued with a strongly personal element that excludes any purely abstract or rationalistic speculation about the ultimate meaning of reality. What concerns him is that part of reality which is identified with the nature of man. The nature of man is, of course, inseparable from nature in the wider sense, but sensibility and feeling, rather than mere reason, are probably the most effective means of penetrating this wider objective realm of being. The thinker concerned with fundamental truths will do well, in Rousseau's view, to concentrate on what is of interest to him, "interest" here being defined not in any narrowly pragmatic or empirical sense but as indicating those matters that appertain to man's original nature. This means that Rousseau finally emerges as a moralist rather than as a traditional metaphysician.

Since reflection on the nature of man involves the ability to distinguish between reality and appearance, between the genuinely original and the merely artificial aspects of existence, the thinker's first task must be to abandon the illusions of opinion for the truths of nature. This explains both the negative, critical aspects of Rousseau's views of modern society and his more positive, constructive efforts to elaborate a philosophy of man. If his interpretation of nature seemed too optimistic to satisfy the demands of contemporary religious orthodoxy, it was also too religious to please the advocates of philosophical skepticism or materialism. Of one thing Rousseau felt quite certain: To ignore or reject the profound moral and spiritual aspects of human existence could have only the most disastrous consequences for the

welfare of humanity. The discovery of truth requires an active renewal of the whole man and a reawakened moral consciousness that acknowledges the full implications of man's situation in the universe; the genuine possibilities of human life cannot be separated from the universal order of which they are a part, and man's ultimate felicity is to feel himself at one with a God-created "system in which all is good."

Like so many of his contemporaries Rousseau considered happiness to be the legitimate goal of human endeavor, but he insisted that "enjoyment" must not be interpreted in a shallow or selfish manner. Happiness consists of being oneself and of existing according to one's own nature, but a nature that has been purified of all extraneous artificial elements. When truly fulfilled, man will experience satisfaction with himself and a sense of being identified with the pure "feeling of existence"; this, in turn, presupposes the ability to find a true personal unity and plenitude. No doubt, Rousseau's efforts to realize this ideal in his own life were not free from ambiguity and contradiction, as an examination of his personal writings well shows, but his didactic works are consistent in their main objective.

In a corrupt society the recovery of a full human existence can never take the form of a mere return to nature, for the nature of man cannot be equated with the primordial state of nature. Although Rousseau was often nostalgically drawn to the innocence and simplicity of early times, he also treated nature as a dynamic, forward-looking concept. Starting from man as he is, the movement toward nature must be constantly sustained by the vision of what man might be. The achievement of this goal requires a radical transformation of human existence, the rediscovery and re-creation of a new nature. At the same time Rousseau did not believe in the need for any kind of supernatural grace to help man to carry out this task, since nature represented an innate possibility that could be realized through the wise exercise of human freedom alone.

Rousseau's powerful influence on later generations was partly due to this vision of a regenerated human nature, but unlike merely utopian thinkers he seemed to promise a transfiguration of everyday existence, not the pursuit of a hopeless chimera. Indeed, his philosophy revealed a striking, if often elusive, combination of idealistic and realistic elements that constantly seemed to open up the possibility of a better world. Moreover, this optimistic outlook was transmitted through a particularly eloquent and persuasive style, rich in emotional and musical overtones, giving the impression of intense sin-

cerity and convincing the humblest of men that he need never feel ashamed to call himself a human being.

**See also** Alembert, Jean Le Rond d'; Analytical Feminism; Authority; Buffon, Georges-Louis Leclerc, Comte de; Condillac, Étienne Bonnot de; Diderot, Denis; Encyclopédie; Equality, Moral and Social; French Philosophy; General Will, The; Human Nature; Hume, David; Malebranche, Nicolas; Philosophy of Education, History of; Plato; Plutarch of Chaeronea.

## Bibliography

### WORKS BY ROUSSEAU

For bibliographical information about the various editions of Rousseau's works, though not about works on Rousseau, the reader is referred to Jean Sénélier, *Bibliographie générale des oeuvres de J.-J. Rousseau* (Paris, 1949), and Théophile Dufour, *Recherches bibliographiques sur les oeuvres imprimées de J.-J. Rousseau*, 2 vols. (Paris: L. Giraud-Badin, 1925).

### Collections and Other Editions

Of the collected editions of Rousseau's works, *Oeuvres complètes de J.-J. Rousseau*, 13 vols. (Paris, 1865–1870), and *Oeuvres et correspondance inédites de J.-J. Rousseau*, edited by G. Streckeisen-Moultou (Paris, 1861), will be superseded by *Oeuvres complètes de Jean-Jacques Rousseau*, edited by Pléiade Bernard Gagnebin and Marcel Raymond (Paris, 1959–), five volumes projected. Similarly, the *Correspondance générale de J.-J. Rousseau*, edited by Théophile Dufour and P.-P. Plan, 20 vols. (Paris: Colin, 1924–1934), with the *Table de la correspondance générale*, edited by P.-P. Plan and Bernard Gagnebin (Geneva, 1953), superseded by the *Correspondance complète de Jean-Jacques Rousseau*, 51 vols., edited by R. A. Leigh (Geneva, 1965–1991).

Although the new Pléiade edition of the collected works will henceforth be authoritative, the following separate editions are still important: *Political Writings*, edited by C. A. Vaughan, 2 vols. (Cambridge, U.K.: Cambridge University Press, 1915); the critical edition of *La nouvelle Héloïse*, edited by Daniel Mornet, 4 vols. (Paris, 1925); *La profession de foi du vicairé savoyard*, edited by P. M. Masson (Paris, 1914); *Les rêveries du promeneur solitaire*, edited by J. S. Spink (Paris, 1948). For the *Confessions* see also the edition by Jacques Voisine (Paris, 1964).

### WORKS ON ROUSSEAU

Some general bibliographical information about studies of Rousseau's work is to be found in Albert Schinz, *État présent des travaux sur J.-J. Rousseau* (New York: Modern Language Association of America, 1941); Jacques Voisine, "État des travaux sur J.-J. Rousseau," in *L'information littéraire* 16 (May–June 1964): 93–102; and Peter Gay, *The Party of Humanity, Studies in the French Enlightenment* (New York: Knopf, 1964), Ch. 8, "Reading about Rousseau."

A very pleasant introduction to Rousseau is Bernard Gagnebin's *A la Rencontre de Jean-Jacques Rousseau* (Geneva, 1962), a picturesque collection of texts, documents, and illustrative material. Sound general

introductions to Rousseau's thought are Daniel Mornet, *Rousseau l'homme et l'oeuvre* (Paris, 1950); E. H. Wright, *The Meaning of Rousseau* (Oxford: Oxford University Press, H. Milford, 1929), which corrects earlier misinterpretations of Rousseau's idea of nature; and J. H. Broome, *Rousseau: A Study of His Thought* (New York: Barnes and Noble, 1963), to which should be added *Rousseau par lui-même*, edited by Georges May (Paris: Éditions du Seuil, 1962), the pioneering article by Gustave Lanson, "L'unité de la pensée de Jean-Jacques Rousseau," *Annales de la société Jean-Jacques Rousseau* 8 (1912), and Ernst Cassirer, *The Question of Jean-Jacques Rousseau*, translated and edited with an introduction by Peter Gay (New York: Columbia University Press, 1954).

For Rousseau's biography see Jean Guéhenno, *Jean-Jacques*, 3 vols. (Paris: Grasset, 1948–1952), of which a new edition, titled *Jean-Jacques, histoire d'une conscience*, 2 vols., was published in 1962. See also F. C. Green, *Jean-Jacques Rousseau: A Critical Study of His Life and Writings* (Cambridge, U.K.: Cambridge University Press, 1955).

More detailed discussions of various aspects of Rousseau's philosophy are to be found in C. W. Hendel, *Jean-Jacques Rousseau: Moralism*, 2 vols. (New York and Oxford: Oxford University Press, 1934; 2nd ed., 1 vol., New York, 1962), an important examination of Rousseau's intellectual development; Albert Schinz, *La pensée de Jean-Jacques Rousseau* (Paris, 1929), which stresses and perhaps exaggerates the basic conflict between the "romantic" and the "Roman" Rousseau but is nevertheless a significant study; Pierre Burgelin, *La philosophie de l'existence de J.-J. Rousseau* (Paris: Presses Universitaires de France, 1952), an important modern synthesis of Rousseau's thought; Robert Derathé, *Le rationalisme de Jean-Jacques Rousseau* (Paris, 1948), a helpful corrective to earlier sentimentalist interpretations of Rousseau; and Jean Starobinski, *Jean-Jacques Rousseau, la transparence et l'obstacle* (Paris: Plon, 1957), an original and important study of certain key themes in Rousseau's work.

Still important for a study of Rousseau's Genevan background is Gaspard Vallette, *Jean-Jacques Rousseau genevois* (Paris, 1908), although the discussion of the Genevan aspects of Rousseau's thought has been modified by more recent criticism, especially by J. S. Spink, *Rousseau et Genève* (Paris, 1934), and François Jost, *Jean-Jacques Rousseau suisse*, 2 vols. (Fribourg: Éditions Universitaires, 1961).

Indispensable for any serious study of Rousseau's religious thought is P. M. Masson's *La religion de J.-J. Rousseau*, 3 vols. (Paris, 1916), in spite of some exaggeration of both its Roman Catholic and sentimental elements; for a corrective see Albert Schinz, *La pensée religieuse de J.-J. Rousseau et ses récents interprètes* (Paris, 1927). On Rousseau's political thought see Alfred Cobban, *Rousseau and the Modern State* (London: Allen and Unwin, 1934; 2nd ed., 1964), and Robert Derathé, *Jean-Jacques Rousseau et la science politique de son temps* (Paris: Vrin, 1950), which sets Rousseau's political thought in its contemporary philosophical context. For the difficult question of Rousseau's psychology and personality see Louis Proal, *La psychologie de J.-J. Rousseau* (Paris, 1930), and Suzanne Elosu, *La maladie de J.-J. Rousseau* (Paris, 1929). More recent discussions of this problem and its bearing on the personal writings are to be found in Ronald Grimsley, *J.-J. Rousseau: A Study in Self-Awareness* (Cardiff: University of Wales Press, 1961), and

Marcel Raymond, *J.-J. Rousseau: La Quête de soi et la rêverie* (Paris: Librairie J. Corti, 1962).

The year 1962, being the 250th anniversary of Rousseau's birth, was marked by three important international conferences whose proceedings were published: *Jean-Jacques Rousseau et son oeuvre, Colloque de Paris, 16–20 Octobre*, 1962 (Paris: Librairie C. Klincksieck, 1964); *Annales de la société Jean-Jacques Rousseau*, Vol. 35, *Entretiens sur J.-J. Rousseau* (Geneva, 1962); and *Études sur le contrat social de J.-J. Rousseau* (Paris: Belles Lettres, 1964), proceedings of Dijon conference of May 1962.

It has not been possible to include in this bibliography many important articles on Rousseau. For further information on this and other subjects, the reader is referred to the indispensable *Annales de la société Jean-Jacques Rousseau* (Geneva), published from 1905 on, which contain not only original articles but a full review of Rousseau literature.

**Ronald Grimsley (1967)**

## ROUSSEAU, JEAN-JACQUES [ADDENDUM]

The writings of Jean-Jacques Rousseau continue to attract a wide range of readers throughout the world. Persistent questions concerning nationalism, political legitimacy, and the social costs of technological progress sustain an ongoing interest in Rousseau's major political writings (*The Social Contract*, *Considerations on the Government of Poland*, the first and second discourses). Controversies over child-rearing, the nature of language, and the role of the media in public life keep alive his educational and cultural writings (*Emile, Essay on the Origin of Languages*, *Letter to d'Alembert on the Theater*). Speculations about psychology and the arts of autobiography draw readers to Rousseau's personal writings (*The Confessions*, *Reveries of a Solitary Walker*, *Rousseau Judge of Jean-Jacques*). And new attitudes regarding love, marriage, and eroticism provoke reconsideration of his romantic novel (*La nouvelle Héloïse*). As the editors of a 1978 issue of *Daedalus* commemorating the bicentennial of Rousseau's death observed, Rousseau anticipated many of the moral, political, social, and aesthetic concerns that continue to preoccupy us today.

Three intellectual currents have contributed significantly to a growing body of scholarship on Rousseau. Feminist studies have offered fresh interpretations of his notoriously controversial writings about the nature, education, and status of women (see esp. *Emile*, book 5). Some feminist theorists (e.g., Okin, 1979) argue that Rousseau's advocacy of sexually differentiated social and political roles contradicts his egalitarian principles and

undermines the logic and validity of his political theory. Others (e.g., Weiss, 1994) maintain that sexual differentiation constitutes a necessary social construct undergirding the unity of his entire system. At issue in many of these debates are fundamental questions about the usefulness for modern feminism of any theory that posits a close connection between a woman's essential "nature" and her moral role in society.

Deconstruction has also affected the content and direction of Rousseau criticism, especially among scholars in language and literature departments. The French philosophers and literary critics who originated this movement in the 1960s and 1970s gave prime place to Rousseau in the development of their ideas (see, e.g., Derrida, 1976). In seeking to expose the indeterminacy of the meaning of Rousseau's texts by examining details that are commonly overlooked (e.g., footnotes, metaphors, his choice of particular terms), deconstructionist critiques illuminate the multilayered quality of his prose and show that even an author committed to the truth may produce writings fraught with artifice.

A third important source of Rousseau criticism has been the legacy of Leo Strauss (1899–1973)—a political philosopher who is as well known for the habits of close textual analysis he passed on to his students as for the ideas put forth in his own writings (see, e.g., Strauss, 1953). Straussian interpretations take seriously Rousseau's claims that his political thought forms a single coherent system; they also emphasize his debt to classical sources. Most important, perhaps, the Straussian legacy includes a substantial number of English translations of Rousseau's work (e.g., by Allan Bloom, Victor Gourevitch, Christopher Kelly, Judith R. Bush, and Roger D. Masters)—thus making him more accessible to the general reader in North America.

Rousseau specialists have benefited from the publication of Rousseau's *Oeuvres complètes* and *Correspondance complète*, from the appearance of scholarly journals and associations devoted to Rousseau studies (*Annales de la Société Jean-Jacques Rousseau*, *Études Jean-Jacques Rousseau*, and the *Proceedings* of the North American Association for the Study of Rousseau), and from the publication of papers delivered at various conferences held in 1978 to commemorate his death and in 1989 to mark his relationship to the French Revolution.

**See also** Deconstruction; Derrida, Jacques; Love; Nationalism; Rousseau, Jean-Jacques.



**Bibliography**

## WORKS BY ROUSSEAU

- Oeuvres complètes*. 5 vols, edited by B. Gagnebin and M. Raymond. Paris: Gallimard, 1959–.
- Correspondance complète de Jean-Jacques Rousseau*. 51 vols, edited by R. A. Leigh. Geneva: Institut et Musée Voltaire, 1965–1991.
- Emile: or, On Education*. Translated by A. Bloom. New York: Basic, 1979.
- The Collected Writings of Jean-Jacques Rousseau*, edited by R. D. Masters and C. Kelly; translated by J. R. Bush, C. Kelly, and R. D. Masters. Hanover, NH, 1990–.

## WORKS ON ROUSSEAU

- Affeldt, Steven G. “The Force of Freedom: Rousseau on Forcing to Be Free.” *Political Theory* 27 (3) (1999): 299–333.
- Bloom, Harold, ed. *Jean-Jacques Rousseau*. New York: Chelsea House, 1988.
- Boyd, Richard. “Pity’s Pathologies Portrayed: Rousseau and the Limits of Democratic Compassion.” *Political Theory* 32 (4) (2004): 519–546.
- Charvet, John. *The Social Problem in the Philosophy of Rousseau*. Cambridge, U.K.: Cambridge University Press, 1974.
- Cooper, Laurence D. *Rousseau, Nature, and the Problem of the Good Life*. University Park: Pennsylvania State University Press, 1999.
- Cranston, Maurice William and Richard S. Peters, eds. *Hobbes and Rousseau: A Collection of Critical Essays*. Garden City, NY: Anchor Books, 1972.
- Cranston, Maurice William. *Jean-Jacques: The Early Life and Work of Jean-Jacques Rousseau, 1712–1754*. New York: Norton, 1982.
- Cranston, Maurice William. *The Noble Savage: Jean-Jacques Rousseau, 1754–1762*. Chicago: University of Chicago Press, 1991.
- Cranston, Maurice William. *The Solitary Self: Jean-Jacques Rousseau in Exile and Adversity*. Chicago: University of Chicago Press, 1997.
- Dascal, Marcelo. “‘Aporia’ and ‘Theoria’: Rousseau on Language and Thought.” *Revue Internationale et Philosophie* 32 (1978): 214–237.
- Dent, N. J. H. “Jean-Jacques Rousseau.” In *The Blackwell Guide to the Modern Philosophers from Descartes to Nietzsche*, edited by Steven Emmanuel. Cambridge, MA: Blackwell, 2001.
- Dent, N. J. H. *Rousseau: An Introduction to His Psychological, Social and Political Theory*. New York: Blackwell, 1988.
- Dent, N. J. H. *A Rousseau Dictionary*. Cambridge, MA: Blackwell Reference, 1992.
- Ferrara, Alessandro. *Modernity and Authenticity: A Study in the Social and Ethical Thought of Jean-Jacques Rousseau*. Albany: State University of New York, 1993.
- Fralin, Richard. *Rousseau and Representation*. New York: Columbia University Press, 1978.
- Grimsley, Ronald. *The Philosophy of Rousseau*. New York: Oxford University Press, 1973.
- Hall, John C. *Rousseau, An Introduction to his Political Philosophy*. London: Macmillan, 1973.
- Havens, George Remington. *Jean-Jacques Rousseau*. Boston: Twayne, 1978.
- Hedman, Carl G. “Rousseau on Self-Interest, Compassion, and Moral Progress.” *Revue de l’université d’Ottawa* 49 (1979): 430–447.
- Hiley, David R. “The Individual and the General Will: Rousseau Reconsidered.” *History of Philosophy Quarterly* 7 (2) (1990): 159–178.
- Hoffman, Piotr. *Freedom, Equality, Power: The Ontological Consequences of the Political Philosophies of Hobbes, Locke, and Rousseau*. New York: Lang, 1999.
- Horowitz, Asher. *Rousseau, Nature, and History*. Toronto: University of Toronto Press, 1987.
- Howard, Dick. “Rousseau and the Origin of Revolution.” *Philosophy and Social Criticism* 6 (1979): 349–370.
- Johnston, Steven. *Encountering Tragedy: Rousseau and the Project of Democratic Order*. Ithaca, NY: Cornell University Press, 1999.
- Kain, Philip J. “Rousseau, the General Will, and Individual Liberty.” *History of Philosophy Quarterly* (1990): 315–334.
- Kavanagh, Thomas M. *Writing the Truth: Authority and Desire in Rousseau*. Berkeley: University of California Press, 1987.
- Kukla, Rebecca. “The Coupling of Human Souls: Rousseau and the Problem of Gender Relations.” In *Political Dialogue: Theories and Practices*, edited by Stephen L. Esquith. Amsterdam: Rodopi, 1996.
- Lange, Lynda. *Feminist Interpretations of Jean-Jacques Rousseau*. University Park: Pennsylvania State University Press, 2002.
- Lauth, Reinhard. “Le système des pensées de Rousseau selon a Philonenko.” *Archives de Philosophie* 50 (1987): 23–53.
- Leigh, R. A., ed. *Rousseau after Two Hundred Years: Proceedings of the Cambridge Bicentennial Colloquium*. New York: Cambridge University Press, 1982.
- Levine, Andrew. *The General Will, Rousseau, Marx, Communism*. New York: Cambridge University Press, 1993.
- Levine, Andrew. “Beyond Justice: Rousseau against Rawls.” *Journal of Chinese Philosophy* 4 (1977): 123–142.
- Masters, Roger D. *The Political Philosophy of Rousseau*. Princeton, NJ: Princeton University Press, 1968.
- Miller, Jim. *Rousseau: Dreamer of Democracy*. New Haven, CT: Yale University Press, 1984.
- Morgenstern, Mira. *Rousseau and the Politics of Ambiguity: Self, Culture, and Society*. University Park: Pennsylvania State University Press, 1996.
- Morris, Christopher W. *The Social Contract Theorists: Critical Essays on Hobbes, Locke, and Rousseau*. Lanham, MD: Rowman and Littlefield, 1999.
- O’Hagan, Timothy, ed. *Jean-Jacques Rousseau and the Sources of the Self*. Brookfield, MA: Avebury, 1997.
- O’Hagan, Timothy. *Rousseau*. London: Routledge, 1999.
- Orwin, Clifford and Nathan Tarcov, eds. *The Legacy of Rousseau*. Chicago: University of Chicago Press, 1997.
- Perkins, Merle L. *Jean-Jacques Rousseau on the Individual and Society*. Lexington: University Press of Kentucky, 1974.
- Qvortrup, Mads. *The Political Philosophy of Jean-Jacques Rousseau: The Impossibility of Reason*. New York: Manchester University Press, 2003.
- Rapaczynski, Andrzej. *Nature and Politics: Liberalism in the Philosophies of Hobbes, Locke, and Rousseau*. Ithaca, NY: Cornell University Press, 1987.

- Reisert, Joseph R. *Jean-Jacques Rousseau: A Friend of Virtue*. Ithaca, NY: Cornell University Press, 2003.
- Riley, Patrick, ed. *The Cambridge Companion to Rousseau*. New York: Cambridge University Press, 2001.
- Rosenblatt, Helena. *Rousseau and Geneva: From the "First Discourse" to the "Social Contract," 1749–1762*. New York: Cambridge University Press, 1997.
- Ryan, Bruce A. "Jean-Jacques Rousseau and Behavior Control: The Technology of a Romantic Behaviorist." *Behaviorism* 4 (1976): 245–256.
- Scholz, Sally J. *On Rousseau*. Belmont, CA: Wadsworth/Thomson Learning, 2001.
- Schwartz, Joel. *The Sexual Politics of Jean-Jacques Rousseau*. Chicago: University of Chicago Press, 1984.
- Shaver, Robert. "Rousseau and Recognition." *Social Theory and Practice* 15 (1989): 261–283.
- Simon Ingram, Julia. "Rousseau and the Problem of Community: Nationalism, Civic Virtue, Totalitarianism." *History of European Ideas* 16 (1–3) (1993): 23–29.
- Starobinski, Jean. *Jean-Jacques Rousseau, Transparency and Obstruction*. Translated by Authur Goldhammer. Chicago: University of Chicago Press, 1988.
- Steinberg, Jules. *Locke, Rousseau, and the Idea of Consent: An Inquiry into the Liberal-Democratic Theory of Political Obligation*. Westport, CT: Greenwood Press, 1978.
- Still, Judith. *Justice and Difference in the Works of Rousseau*. New York: Cambridge University Press, 1993.
- Strong, Tracy B. *Jean Jacques Rousseau: The Politics of the Ordinary*. Lanham, MD: Rowman and Littlefield, 2002.
- Trouille, Mary Seidman. *Sexual Politics in the Enlightenment: Women Writers Read Rousseau*. Albany: State University of New York Press, 1997.
- Viroli, Maurizio. *Jean-Jacques Rousseau and the 'Well-Ordered Society'*. Translated by Derek Hanson. New York: Cambridge University Press, 1988.
- Vulliamy, Colwyn E. *Rousseau*. Port Washington, NY: Kennikat Press, 1972.
- Weirich, Paul. "Rousseau on Proportional Majority Rule." *Philosophy and Phenomenological Research* 47 (1986): 111–126.
- Weiss, Penny A. *Gendered Community: Rousseau, Sex, and Politics*. New York: New York University Press, 1993.
- Wokler, Robert. *Rousseau*. New York: Oxford University Press, 1995.
- Wokler, Robert. *Rousseau: A Very Short Introduction*. New York: Oxford University Press, 2001.

**Grace G. Roosevelt (1967)**

*Bibliography updated by Tamra Frei (2005)*

## ROYCE, JOSIAH (1855–1916)

Josiah Royce, the American idealist philosopher, was born in Grass Valley, California. He received his AB degree from the University of California in 1875 and his doctorate from Johns Hopkins University in 1878. In the inter-

vening years he studied in Germany at Leipzig and Göttingen, where he attended the lectures of Hermann Lotze. Royce returned to the University of California in 1878 as an instructor of English. Four years later, with the help of William James and George Herbert Palmer of the Harvard department of philosophy, he was invited to Harvard, where he taught for two years as a replacement for men on leave; in 1885 he received a regular appointment as assistant professor. Until his death Royce was one of the mainstays of the philosophy department in its so-called golden period. During that time he carried on his friendly debate with William James about the merits and demerits of absolute idealism, supervised the doctoral work of George Santayana, and delivered the Gifford Lectures at the University of Aberdeen in Scotland. Royce was a prolific writer and was much in demand as a public speaker.

### PHILOSOPHICAL ORIENTATION

Royce's philosophy is a unique synthesis of the rationalist metaphysic we associate with the system builders in the Western philosophical tradition and the appeal to experience and practice that has been dominant in American philosophy since 1875. Royce is the best American representative of absolute idealism, although there are voluntaristic elements in his position that distinguish it from both the Hegelian position and the systems of the British idealists. Royce's theory of the will and his conception of its role in the knowledge process introduced novel features into the tradition of rationalistic idealism. Royce was aware of this fact and hence called his position absolute voluntarism or absolute pragmatism.

Royce's thought revolves around the problems raised by a religious view of reality. He sought to resolve them through a metaphysical system constructed with the aid of concepts drawn from a wide range of thought and experience. Basic to his position is the concept of the self, an idea that he elucidated in several forms. In his earlier thought the self appears as the Absolute Knower, grasping all truth in one synoptic vision *totum simul*. Later, however, Royce put more emphasis on mediation and on the idea of system. Ultimately, he arrived at the community of interpretation, or social theory of reality, according to which all selves are joined in a Universal Community whose goal is to possess the truth in its totality.

### THE NATURE OF BEING

In large measure Royce's idealism consists in his having given to the process of knowing a privileged position in the definition of reality. The nature of Being is to be

determined through the elucidation of the process of being known.

**ARGUMENT FROM ERROR.** The pattern of the approach through knowing was established early in Royce's development. In a paper, "Kant's Relation to Modern Philosophic Progress" (1881), he argued that the proper task of philosophy is to study the nature of experience, especially the role played by the forms of intellectual activity in knowing. In later works he returned repeatedly to the task of defining the relation between sense and understanding, between the perceptual and conceptual poles in experience and knowledge. Strongly influenced by Immanuel Kant, Royce sought to discover the exact relation between the knowing activity and its matter. He asked how the function of judgment transforms the sensible starting point of all experience into knowledge. Whereas Kant had argued that the past moment and its datum can be brought into the present through the activity of the transcendental subject, Royce regarded the past and future as projections from the present. Knowledge starts with immediate data of sense; these data, as present, are beyond the control of judgment (this is the realistic element in Royce's idealism), but the whole of experience involving reference to a past, a future, and a public object is to be built up from the momentary consciousness. In order to accomplish this construction, judgment and principles of transcendence are required.

Dissatisfied with the view that assigns the status of postulates to the principles needed for transforming immediate data into knowledge, Royce sought to justify those principles. His theory of the Absolute Knower, which he developed in the well-known chapter "The Possibility of Error" in *The Religious Aspect of Philosophy* (Boston, 1885), was intended to show that the conditions for both knowledge and error must themselves be actual; what is actual cannot be explained or justified by what is merely possible or postulated. The argument that is presented for the existence of God or the Absolute Knower may be summarized as follows. Error actually exists; erroneous judgments cannot be made erroneous by finite knowers. In order to be in error, a judgment must fail to agree with its intended object. Yet if the intended object is wholly and completely defined by the isolated judgment, it is difficult to see how the judgment can fail to be true. Royce's central contention is that a judgment can have its own object and at the same time fail to agree with it only if the judgment is not isolated as an entirely enclosed fact but is, instead, part of a system of judgments or an organized body of thought. The isolated judgment cannot have within itself the distinction between its truth and falsity;

for that we need an inclusive thought capable of relating the isolated judgment to all other actual and possible judgments about the intended object. In finding error as a fact that we cannot create, we are actually involved in the Infinite Thought. Without that Thought, error is either impossible or unintelligible. This ingenious argument assumes, among other things, that the real individual at which knowledge aims can be identified only at the end of the knowledge process. However, as Charles Peirce and others have shown, there is no need to make this assumption, although without it the argument fails.

**THOUGHT AND REALITY.** Royce continued to approach the problem of Being—the problem of defining the basic nature of the real—through concentration on the knowledge process. He was also trying to retain critical philosophy and neutralize its negative judgment on the possibility of ontology. His solution was to say that a theory of Being is possible if we can discover the true relation between our ideas and the real world. In *The World and the Individual* (New York, 1901–1902) Royce posed the problem of Being as one of explaining what thought and reality must be like if the former is to attain genuine knowledge of the latter. By means of an extended dialectical argument, Royce examined three classical theories of Being (in his language, theories of "the ontological predicate")—realism, mysticism, and critical rationalism. In subjecting them to critical analysis, he tried to show the element of truth and error in each. From this analysis Royce's own voluntaristic idealism emerged; it was designed to avoid the errors of the other positions while preserving their truth in a new and more comprehensive system that defined Being in terms of purpose fulfilled.

For Royce realism is the doctrine that to be is to be independent of being known. According to realism, the real is just what it is apart from the knower and his acts of knowledge. Royce, however, aimed at exposing this position and hence placed a narrow construction on the term *independent*. To be independent is taken to mean that the idea and object are totally externally related. If the idea and object are thus disconnected, he argued, then knowledge becomes inexplicable, and reality is severed from truth. Peirce, among others, objected to this statement of the realist position, describing it as one-sided.

Mysticism is defined as the thesis that to be is to be immediate. Here again, the real is understood as that which falls effectively beyond the power of analytical reason.

Royce's exposition of critical rationalism, which he defined somewhat cryptically with the formula "to be is to be valid," has been charged with ambiguity; John Dewey claimed that Royce's entire argument was vitiated by his having confused "possible experience" and "validity" in his presentation of the position. Dewey's claim is not without warrant; Royce combined several ideas under one heading, and it is not clear that they are compatible. Nevertheless, Royce's argument is clear enough in its main outline. The critical rationalist does not accept the independent objects of either realism or common sense and still less allows the immediacy of mysticism. Instead, he defines the real as that which gives warrant or validity to our ideas. To be real in this instance means that an object conforms to certain universal forms or conditions—causal sequence, temporal succession, spatial relations, numerical identity, and so on—that are marked out in advance as the general structure of all experience. For Royce the merit of this position is that it comes closer to defining reality in terms of truth than was possible with either realism or mysticism. Critical rationalism, however, is inadequate because it can define or anticipate only the universal form of experience and cannot reach the determinate individual. Royce's point is that the determinate individual cannot be defined in terms of universal conditions of possible experience alone; in order to have knowledge of an individual, we must appeal to actual, sensible experience. But it is just the need for this appeal that marks the defect of the position; a completed rationalistic idealism would show us how to pass from the idea to its fulfillment in the individual object without having to appeal to a brute, sensible experience that is "given." Critical rationalism, however, is forced to rest with "possible experience," by which Royce meant the universal conditions that any proposed object of knowledge would have to satisfy in order to be an object of experience at all. It is important to notice that the entire discussion is dialectical, in the sense that Royce expounds and criticizes the alternative theories only in relation to his own final view. Competing theories fail or succeed precisely to the extent that they are incompatible with, or contribute to the development of, his voluntaristic idealism.

**VOLUNTARISTIC IDEALISM.** Royce's own view can be summed up in the thesis that to be is to be the individual or determinate fulfillment of a purpose. Distinguishing between the internal and external meaning of ideas, Royce defined an idea as a purpose (internal meaning) seeking its object, or other (external meaning). An idea intends, and thus selects, its object; the object, as the full realization of the idea, must be the determinate individ-

ual that allows no other of its kind if it is to be the unique fulfillment of the purpose expressed by the original idea. If we say that Socrates is snub-nosed, our ideas (internal meaning) aim at, or intend, the unique and unduplicable individual Socrates (external meaning). Our ideas are not about just anyone or anything but only about the individual intended; the internal meaning selects the object (external meaning) by reference to which it can be judged true or false. The voluntarism of the position lies in the idea that the other at which all ideas aim is itself the expression of the absolute will or purpose. For Royce it is only in this way that we can explain how an idea can correspond with an object other than itself while that object remains other and yet is the object intended by the idea.

The entire theory is recognizable as a modern version of an ancient doctrine of self-knowledge. We start with an idea that is fragmentary and imperfectly understood, and we seek to find its true meaning in the object that is its individual fulfillment. The object intended exceeds the fragment with which we began; we can discover the true nature of the object and the truth or falsity of our idea only when we have reached the total individual reality that fulfills our purpose. Royce developed this conception of Being into a comprehensive system embracing a doctrine of man, nature, and God. The rational will and its purpose mark the ultimate reality; all finite individuality is what it is in virtue of its fulfilling the purpose of the Absolute Self.

*The reality of the infinite.* In the essay "The One, the Many and the Infinite" appended to *The World and the Individual*, Royce introduced the topic that was to occupy much of his later thought—the reality of the infinite. He attempted to refute the claim, made by F. H. Bradley in *Appearance and Reality* (1893), that we cannot express in clear concepts the detail of the many facts constituting the Absolute. Since such a claim, if true, would have rendered Royce's entire project pointless, he felt called upon to refute it. To explain how the many develop out of the one, Bradley argued, always leads to an actual infinity, and this is self-contradictory. In the Absolute all is one, but according to Bradley, we are unable to comprehend the unity. Royce denied that an actual infinite is self-contradictory. Through the concept of a self-representative system based on what would now be called a recursive function, he developed a modern version of the actual infinite. The form of the self-representative system was construed as a purpose or an ordering plan and defines once and for all an actual infinity of members. A self-representative system is one that represents itself with all else that it represents. A mirror of the entire universe, for

example, would have to include itself among the represented items. By the form of the system, Royce meant the principle or purpose behind it, which in the above example would be mirroring. From the one form or purpose there comes, by the recurrent or self-representative operation, an infinity of detail such that nothing less than that infinity will serve to express all that was meant by the original form. Understanding the self as having the form of a self-representative system, Royce claimed that the multitude of details constituting the concrete individuality of the real world is an expression of that self. Reality is an actual infinite, a unity of one and many. Royce's later doctrine of the community of interpretation represents his final attempt to elaborate the theory.

*Logic and mathematics.* It is important to note that Royce took very seriously the development of mathematical logic and studies in the foundation of mathematics. He was fond of criticizing pragmatism for neglect of what he took to be a doctrine of absolute truth implied in the new logic of Gottlob Frege, Bertrand Russell, Giuseppe Peano, and Ernst Schröder. Maintaining that "order is the fundamental category of exact thought about facts," Royce argued for the validity of using technical logical concepts in the construction of a metaphysical theory. Two examples will clarify the point. In the analysis of discrimination, he used the concept of between, arguing that discrimination and comparison are possible because, given any two conceptions, we are always able to find a third conception that is between the other two and expresses some relation in which they stand. This point was later expressed through the logic of triadic relations and the theory of interpretation. An even more striking illustration is found in the use of the limit concept to define the nature of the real as individual. The reality at which the process of knowledge is directed is said to be the "limit" of a series of attempts to apprehend the object. Royce understood "limit," not in the sense of an end term that we can approach at will, but in the sense of a least upper bound, which, in the series  $1 + \frac{1}{2} + \frac{1}{4} \dots$ , for example, is the least number that lies beyond the sum of the series—namely, 2. Thus, the real, individual reality is what is immediately beyond the whole series of efforts to know it.

## ETHICAL AND RELIGIOUS DOCTRINES

Royce contributed ideas worthy of consideration to almost every branch of philosophy, not least in ethics.

**LOYALTY.** Royce's *Philosophy of Loyalty* (New York, 1908) is still one of his best-known books. In it he developed the

principle of loyalty to loyalty as the basic moral law. He regarded his principle as superior to both Kant's categorical imperative and J. S. Mill's principle of utility. Loyalty, by which is meant a freely chosen and practical devotion to a cause or goal, is the highest virtue. Royce was well aware of the existence of evil causes and of the fact that not every cause aims at the loyal spirit. Hence, he argued that loyalty in the ethical sense means devotion to causes that extend the spirit of loyalty and do not contribute to deception, dishonesty, racial and social strife, and so on. Every cause involves some loyalty, but not all causes involve loyalty to loyalty. It is only through loyalty to loyalty itself, the virtue that makes all social life possible, that the self can solve the basic problem of ethics, which is to find a good that is at once objective, in the sense that it constrains our purely individual and subjective interests, and freely chosen, so that the self can acknowledge its obligatory character. Royce followed G. W. F. Hegel in finding the good in a form of self-realization, and he followed Kant in upholding the autonomy of the will.

**PHILOSOPHY OF RELIGION.** Royce's interest in the philosophy of religion was a basic factor in the shaping of his philosophical position. Religious issues constitute the foundation of his thought, starting with *The Religious Aspect of Philosophy* (Boston, 1885) and continuing to his last major work, *The Problem of Christianity* (New York, 1913). Royce had a twofold aim in the philosophical treatment of religion. First, he sought to reinterpret classical religious ideas through contemporary experience and current language; second, he attempted to assess their validity by comparing them with the results of metaphysical analysis. Both aims are clearly present in *The Problem of Christianity*, in which he developed an original interpretation of the Christian religion, first, by uncovering the experiential roots of three central ideas—the church, sin, and atonement—and, second, by seeking support for these ideas in his metaphysics of interpretation and community.

Starting with the view that neither perception nor conception alone, nor any indeterminate combination of the two, is able to yield knowledge of selves, Royce went on to develop the theory of interpretation, according to which all our knowledge is mediated through signs. From this view it follows that the human self is not known (either by itself or another) intuitively as a particular datum or as a universal character but only as the goal of an infinite process of interpretation. In requiring comparison with other selves, this process necessitates a community if there is to be self-knowledge. Persons are involved in, and linked together by, a number of different

communities—political, legal, economic, moral, religious—each of which is defined by its purpose or the goal for which it exists. The religious or Beloved Community has the special purpose of redeeming man from sin (a moral burden) and from the consequences of the self-centered deeds by which he endangers the community through disloyalty. The three central ideas of Christianity (the church, sin, and atonement) are linked together. The Beloved Community is the locus of the love (in Royce's terms, loyalty) exemplified by the atoning deed of Jesus; the church exists to overcome, through love, the self-centeredness of the individual and to transmute the evil consequences of treachery by a constant renewal of the community of many selves devoted to the cause of charity.

The novel feature of Royce's reinterpretation of Christianity is his attempt to rework the much neglected doctrine of the Spirit, or Third Person, of the ancient Trinitarian tradition. God now appears as the Spirit or Interpreter, linking together a multiplicity of distinct selves in a spiritual unity of love. The Beloved Community, founded by the sacrificial or atoning deed of Jesus, becomes the ultimate instrument of the redemptive process.

Unlike William James, Royce was clearly dissatisfied with a purely practical basis for religious belief. Instead, he made the validity of religion dependent on a metaphysical system. He set forth one such system in *The World and the Individual*, and he returned to the task in *The Problem of Christianity*, in which he dealt with specifically Christian ideas. In the intervening years Royce fell under the influence of Peirce's thought, and he freely acknowledged an indebtedness to Peirce's theory of signs, his analysis of triadic relations, and the idea of the community of knowers engaged in interpreting the meaning of things through an infinite system of signs.

The continuation of the logical and epistemological aspects of Royce's philosophy is to be found mainly in the work of C. I. Lewis, and its metaphysical aspects are developed in the thought of W. E. Hocking. The strong current of pragmatism on the American scene, however, carried philosophical thinking away from the speculative realm and directed it into other channels.

**See also** Absolute, The; American Philosophy; Being; Bradley, Francis Herbert; Dewey, John; Frege, Gottlob; Hegel, Georg Wilhelm Friedrich; Hocking, William Ernest; Idealism; James, William; Kant, Immanuel; Lewis, Clarence Irving; Lotze, Rudolf Hermann; Loyalty; Mathematics, Foundations of; Mill, John Stuart;

Mysticism, Nature and Assessment of; Peano, Giuseppe; Peirce, Charles Sanders; Pragmatism; Realism; Relations, Internal and External; Russell, Bertrand Arthur William; Santayana, George; Self-Knowledge.

## Bibliography

### WORKS BY ROYCE

- The Basic Writings*. Edited by John J. McDermott. Chicago: University of Chicago Press 1969.
- The Conception of God: A Philosophical Discussion concerning the Nature of the Divine Idea as a Demonstrable Reality*. St. Clair Shores, MI: Scholarly Press, 1971.
- The Conception of Immortality*. New York: Greenwood Press, 1968.
- Lectures on Modern Idealism*. New Haven, CT: Yale University Press 1967.
- The Letters of Josiah Royce*. Edited by John Clendenning. Chicago: University of Chicago Press 1970.
- The Philosophy of Loyalty*. Nashville, TN: Vanderbilt University Press, 1995.

### WORKS ON ROYCE

- The Problem of Christianity: Lectures Delivered at the Lowell Institute in Boston, and at Manchester College, Oxford*. Hamden, CT: Archon Books, 1967.
- Race Questions, Provincialism, and Other American Problems*. Freeport, NY: Books for Libraries Press 1967.
- The Religious Aspect of Philosophy: A Critique of the Bases of Conduct and of Faith*. Gloucester, MA: Peter Smith, 1965.
- The Religious Philosophy of Josiah Royce*. Westport, CT: Greenwood Press, 1976.
- The Sources of Religious Insight*. New York: Octagon Books, 1977.
- The Spirit of Modern Philosophy: An Essay in the Form of Lectures*. Norton, 1967.
- William James and Other Essays on the Philosophy of Life*. Freeport, NY: Books for Libraries Press 1969.
- For a complete list of Royce's writings, published and unpublished, it is necessary to combine the bibliographies in the following works:
- Cotton, J. H. *Royce on the Human Self*, 305–311. Cambridge, MA: Harvard University Press, 1954.
- Loewenberg, Jacob. *Philosophical Review* 26 (1917): 578–582.
- Rand, Benjamin. *Philosophical Review* 25 (1916): 515–522.
- Smith, J. E. *Royce's Social Infinite*, 171–173. New York: Liberal Arts Press, 1950.

### WORKS ON ROYCE

- Cady, Linell Elizabeth. *The Conceptions of Love and the Self in the Thought of Soren Kierkegaard and Josiah Royce: A Thesis*. PhD diss. Cambridge, MA: Harvard, 1981.
- Clendenning, John. *The Life and Thought of Josiah Royce*. Madison: University of Wisconsin Press, 1985.
- Costello, Harry Todd, and Grover Cleveland Smith. *Josiah Royce's Seminar, 1913–1914: As Recorded in the Notebooks of Harry T. Costello*. Westport, CT: Greenwood Press, 1981.
- Cotton, James Henry. *Royce on the Human Self*. Cambridge, MA: Harvard University Press, 1954; New York: Greenwood Press, 1968. Well-focused; based on manuscript as well as

printed works; aimed at interpreting the whole of Royce's thought in terms of his theory of the self.

- Cunningham, G. Watts. *The Idealistic Argument in Recent British and American Philosophy*. London and New York: Century, 1933. A clear and helpful account of Royce's thought, focusing on his attempt to establish the truth of idealism by the argument from error.
- Fuss, Peter. *The Moral Philosophy of Josiah Royce*. Cambridge, MA: Harvard University Press, 1965. The most incisive account of Royce's ethical philosophy in relation to both his theory of knowledge and the doctrine of the community.
- Galligan, Michael W. E. *God and Experience: A Study of Josiah Royce's "The World and the Individual."* PhD diss. New Haven, CT: Yale University, 1973.
- Imbelli, Robert Peter. *Man's Quest for Salvation in the Thought of Josiah Royce*. PhD diss. New Haven, CT: Yale University, 1973.
- Jarvis, Edward A. *The Conception of God in the Later Royce*. The Hague: Martinus Nijhoff, 1975.
- Kegley, Jacquelyn Ann Kovacevic. *Josiah Royce's Theory of Knowledge*. New York: 1971.
- Kuklick, Bruce. *Josiah Royce: An Intellectual Biography*. Indianapolis: Bobbs-Merrill, 1972.
- Mahoney, James Michael. *Josiah Royce: Moral Reasoning*. PhD diss. Notre Dame, IN: University of Notre Dame, 1975.
- Mahowald, Mary Briody. *An Idealistic Pragmatism: The Development of the Pragmatic Element in the Philosophy of Josiah Royce*. The Hague: Nijhoff, 1972.
- Marcel, Gabriel. *La métaphysique de Royce*. Paris: Aubier, 1945. Translated by V. Ringer and G. Ringer as *Royce's Metaphysics*. Chicago: Regnery, 1956. Four articles on Royce's thought that originally appeared in 1918–1919. The entire range of Royce's philosophy is considered, with indications of Marcel's own existentialism and the resources for it that are to be found in Royce's theory of knowledge and the individual.
- Marcel, Gabriel. *Royce's Metaphysics*. Westport, CT: Greenwood Press, 1975.
- Muirhead, John H. *The Platonic Tradition in Anglo-Saxon Philosophy*. London: n.p., 1931. A fairly complete, but not especially penetrating, account of all of Royce's works from the earliest essays to *The Problem of Christianity*.
- Oppenheim, Frank M. *Bibliography of the Published Works of Josiah Royce*, 1967.
- Oppenheim, Frank M. *Royce's Mature Ethics*. Notre Dame, IN: University of Notre Dame Press, 1993.
- Oppenheim, Frank M. *Royce's Mature Philosophy of Religion*. Notre Dame, IN: University of Notre Dame Press, 1987.
- Powell, Thomas F. *Josiah Royce*. New York: Washington Square Press, 1967.
- Skinner, John E. *The Logocentric Predicament: An Essay on the Problem of Error in the Philosophy of Josiah Royce*. Philadelphia: University of Pennsylvania Press, 1965.
- Smith, John E. *Royce's Social Infinite*. New York: Liberal Arts Press, 1950. A study of Royce's later thought, stressing the idea of the community and of its relation to Peirce's theory of signs. Considerable emphasis is placed on the logic of interpretation and its use in Royce's final statement of the religious issues.

**John E. Smith (1967)**

*Bibliography updated by Michael J. Farmer (2005)*

## ROYER-COLLARD, PIERRE PAUL

(1762–1845)

Pierre Paul Royer-Collard, the French statesman and professor of philosophy, was born at Sompuis, a village in what is now the department of the Marne. He represented this department in the Chamber of Deputies from 1815 to 1839, usually in the opposition. He is best known as the leader of the *Doctrinaires*, a group whose members derived their political views from what they believed to be immutable and self-evident principles. These principles led to a compromise between absolute and constitutional monarchy, and though the principles were supported by Louis XVIII, they were rejected by his brother and successor, Charles X.

Royer-Collard had little, if any, philosophical training. Nevertheless, from 1811 to 1814 he was professor of philosophy and dean at the Sorbonne. He lectured first on Thomas Reid and later on his own views. Just as his political views were a compromise, so in philosophy he sought a compromise between the left wing of sensationalism and the right wing of authoritarian traditionalism. He found it in the philosophy of Reid. Royer-Collard rejected sensationalism on the ground that it could not account for judgment, which is always something contributed to sensory material by the active mind. Since the individual mind is active and capable of making judgments, there is no need of a supernatural authority to dictate to it. In place of such an authority he substituted common sense, which is a consolidation of the judgments of all men. But this did not imply a return to tradition except insofar as tradition itself is an expression of common sense. On the contrary, every man has within him the ability to distinguish between right and wrong, truth and falsity, by a power that resembles the natural light of medieval philosophy. If this faculty did not exist, he maintained, one would be stranded in solipsism, for there would be no reason to believe that one man's conclusions would be harmonious with another's.

Common sense, however, does not operate entirely without the guidance of reason. In reaching its decisions, reason uses two principles of argument, that of causality and that of induction. The search for causes is intrinsic to thinking itself and will inevitably lead back to the idea of a First Cause. For, following Isaac Newton, Royer-Collard believed that one must never accept more causes than are necessary to explain phenomena. However, he does not

seem to have had any clear idea of the nature of a causal explanation.

The principle of induction is a necessary accompaniment to that of causality, for it is by induction that one discovers the essential similarities among phenomena that permit one to group them in a single class. It is man's nature to look for these similarities, as it is his nature to look for causes.

Following Reid, Royer-Collard maintained that the distinction between sensation and perception is all-important. Sensation is simply the pleasure found in experience and is purely subjective. Perception is the apprehension of an external object as external. The externality of the object is not proved by reasoning; it is judged by a spontaneous act of the human mind, as in the twentieth-century epistemology of G. E. Moore.

Though only fragments of Royer-Collard's philosophy exist, collected by his admirer Théodore Jouffroy, it is probable that he saw the philosophy of common sense as a support for his political views. Common sense is the basis of communal life; it provides stable theses of morality and religion; it has all the authority of natural law; and to those who accept it, it is incontrovertible. It is, however, generally admitted that the main contribution of Royer-Collard to French philosophy was the introduction into France of Scottish philosophy.

**See also** Common Sense; Induction; Jouffroy, Théodore Simon; Medieval Philosophy; Moore, George Edward; Newton, Isaac; Reid, Thomas; Sensationalism; Solipsism; Traditionalism.

## Bibliography

### WORKS BY ROYER-COLLARD

*Les fragments philosophiques de Royer-Collard*, edited by André Schimberg. Paris, 1913.

*Discours prononcé à l'ouverture du cours de l'histoire de la philosophie*. A very rare pamphlet in the Bibliothèque de l'Institut.

### WORKS ON ROYER-COLLARD

Boas, George. *French Philosophies of the Romantic Period*, 157–164. Baltimore: Johns Hopkins Press, 1925.

Garnier, Adolphe. "Royer-Collard." In *Dictionnaire des sciences philosophiques*, edited by Adolphe Franck. Paris, 1875.

Spuller, Eugène. *Royer-Collard*. Paris: Hachette, 1895.

Taine, Hippolyte. *Les philosophes classiques de XIX<sup>e</sup> siècle en France*. Paris, 1857. To be read with caution.

**George Boas (1967)**

## ROZANOV, VASILII VASIL'EVICH

(1856–1919)

Vasilii Vasil'evich Rozanov, the Russian critic and philosopher, was born in Vetluga, Russia, and attended secondary schools in Simbirsk and Novgorod before entering Moscow University as a student in the faculty of history and philology. After his graduation from the university in 1881, he taught history and geography in a succession of secondary schools in provincial towns and began the writing on religious and philosophical themes that was to gain him a reputation as a brilliant if erratic critic of contemporary culture, both secular and religious. In 1893 a minor government post in St. Petersburg brought him to the center of Russian literary life, and in 1899 he retired to devote full time to writing. He published numerous books and contributed many articles to the Russian reviews of the day, particularly the reactionary *Novoe vremia* (New times). During the Russian Revolution he took refuge with the religious philosopher Father Pavel Florenskii in Sergiev Posad, near Moscow, where he died.

Rozanov's first major writing and his only strictly philosophical work was an elaborate scholarly treatise titled *O ponimanii* (On the understanding), in which he developed a conception of understanding as a unifying mode of cognition that reconciles science and philosophy. He first won public acclaim with his critical study of Fëdor Dostoevsky, *Legenda o Velikom Inkvizitore* (The legend of the grand inquisitor). In a number of impressionistic, aphoristic works written from 1911 to 1918 he developed most fully the critique of Christianity and the "metaphysics of sex" for which he is best remembered. Chief among these later works are *Opavshie list'ia* (Fallen leaves), *Uedinennoe* (Solitaria), and *Apokalipsis nashego vremeni* (The apocalypse of our time).

Rozanov's mature worldview was a mystical theism based on the sanctification of sex. Emphasizing the generative power of sexuality, Rozanov saw in it the aspect of man that relates him most intimately to God. Sexuality is man's "noumenal aspect," of which his other qualities and capacities are manifestations. Rozanov vigorously attacked Christianity for its denial of the flesh in preaching celibacy and fasting and for its failure to recognize the holiness of elementary animal processes. He preferred the religion of the Old Testament because of what he regarded as its greater acceptance of life and greater humanitarianism, and he called for renewed worship of the vital biological forces enfeebled by Christianity.



**See also** Dostoevsky, Fyodor Mikhailovich; Florenskii, Pavel Aleksandrovich; Mysticism, Nature and Assessment of; Philosophy of Sex; Russian Philosophy.

### Bibliography

#### WORKS BY ROZANOV

*O ponimaniï*. Moscow: Tip. E. Lissiera i IU. Romana, 1886.

*Legenda o Velikom Inkvizitore*. St. Petersburg, 1894.

*Religiia i kul'tura* (Religion and Culture). St. Petersburg, 1899.

*Priroda i istoriia* (Nature and History). St. Petersburg, 1900.

*Temnyi lik: metafizika khristianstva*. St. Petersburg, 1911.

*Uedinennoe*. St. Petersburg, 1912. Translated by S. S.

Koteliansky as *Solitaria*. London: Wishart, 1927. The English edition contains an abridged account of Rozanov's life by E. Gollerbach, other biographical material, and matter from "The Apocalypse of Our Time."

*Opavshie list'ia*. St. Petersburg, 1913. Translated by S. S.

Koteliansky as *Fallen Leaves*. London: Mandrake Press, 1929.

*Izbrannoe* (Selections). Edited by George Ivask. New York, 1956.

#### WORKS ON ROZANOV

Edie, James M., James P. Scanlan, and Mary-Barbara Zeldin, eds. *Russian Philosophy*, 3 vols. Chicago: Quadrangle, 1965.

Poggioli, Renato. *Rozanov*. New York: Hillary House, 1962.

Zenkovsky, V. V. *Istoriia russkoi filosofii*, 2 vols. Paris, 1948–1950. Translated by George L. Kline as *A History of Russian Philosophy*, 2 vols. New York: Columbia University Press, 1953.

*James P. Scanlan (1967)*

## ROZANOV, VASILII VASIL'EVICH [ADDENDUM]

Since the late Soviet period and especially since the collapse of the USSR, Vasilii Rozanov has been one of the most popular and influential thinkers of the religious-philosophical movement of the early twentieth century. His playful and artful texts, at once pungent and profoundly self-conscious, have provided an example of Russian postmodernism *avant là lettre*. He has had a noticeable impact on recent Russian literature (most notably Andrei Siniavskii [aka Abram Terts], Venedikt Erofeev, and Viktor Erofeev) and philosophy (Vladimir Bibikhin). A collection of Rozanov's works, numbering eighteen volumes by the end of 2004, includes many of his published books, scores of uncollected essays, and a wealth of previously unpublished material, including several volumes in the genre of "fallen leaves" (*Sakharna*, 1913; *Mimoletnoe* [Transitory things], 1914 and 1915; *Poslednie list'ia* [Final leaves], 1916 and 1917) and the greater part of texts written for *Apokalipsis nashego vre-*

*meni* (The apocalypse of our time), left unpublished at his death.

Rozanov's flouting of conventional philosophical methods and genres has generated challenging insights in the face of rapidly changing circumstances and intellectual currents. In particular, his phenomenological observations and reflective self-analysis provide an incisive commentary on the interaction between the self and the modern world in such areas as nature, women's fashion, sex, Christianity, and Judaism. In his essay "O sladchaishe m Iisuse i gor'kikh plodakh mira" (On sweet Jesus and the bitter fruits of the world, 1907) Rozanov scandalized Russian orthodoxy by condemning its denial of worldly values. In *Ital'ianskie vpechatleniia* (Italian impressions, 1909) Rozanov used the persona of an estranged and bemused Russian tourist to analyze the legacy of Roman antiquity and the place of religion in modern Western civilization. Even his first book *O ponimaniï* (On understanding), a spectacular flop upon publication in 1886, has been revisited for its more systematic exposition of ideas that Rozanov later developed in his prolific journalistic work.

There has been considerable interest in Rozanov's Aristotelian concepts of potentiality and teleology, and his quest for a holistic form of knowledge more closely attuned to living reality. For Rozanov, cognition cannot be strictly separated from the reality it constructs; long before existentialism, Rozanov declared existence to be prior to essence. He maintains that the philosopher's task is to describe the interaction between the active percipient and his object in all of its complex existence; only in this way can one access and activate essences. Of especial interest is Rozanov's insistence that a particular kind of sympathetic attention is needed to unleash the potential of the inert forms of life. These themes—reality as dependent upon human interaction and history as the process of human understanding—associate Rozanov with the hermeneutic tradition in European philosophy and inspired his turn to an original, narrative style of philosophy in his books of the 1910s. The significance of Rozanov's thought for philosophical aesthetics and the philosophy of religion has only recently begun to be explored.

Rozanov's writings on sexuality (including homosexuality) and Judaism have become only more controversial with time. While Rozanov verged on outright antisemitism in his political commentary, he also professed profound admiration for Jewish traditions of kinship, coupled (especially in his final writings) with increased hostility toward Christian asceticism. His veneration of

fertility caused him to oppose homosexuality, although his frank discussion of the matter rankled with traditionalist allies. Some attempt has been made to cast these writings as “carnivalistic” and “dialogic” (both terms stem from the discourse philosophy of Mikhail Bakhtin): Rozanov is seen as exploring various voices and inverting social conventions in order to overcome the manifest conflicts in society at the level of discourse. Continuing this theme of dialogism, Rozanov’s correspondence (which he often cited in his essays) and his broader intellectual exchange with such leading figures as Konstantin Leont’ev and Pavel Florenskii shed much light on the broader philosophical discourse of his day. As new texts come to light and his entire project is more fully understood, interest in Rozanov’s work will likely continue to grow.

**See also** Aesthetics, History of; Aristotelianism; Bakhtin, Mikhail Mikhailovich; Existentialism; Florenskii, Pavel Aleksandrovich; Hermeneutics; Leont’ev, Konstantin Nikolaevich; Possibility; Philosophy of Religion; Teleology.

## Bibliography

### WORKS BY ROZANOV IN RUSSIAN

- Apokalipsis nashego vremeni*. Moscow: Respublika, 2004.
- Kogda nachal'stvo ushlo*, edited by A. N. Nikoliukin. Moscow: Respublika, 1997.
- Literaturnye izgnanniki: Vospominaniia. Pis'ma*, edited by G. R. Evgrafov. Moscow: Agraf, 2000.
- Mimoletnoe*, edited by A. N. Nikoliukin. Moscow: Respublika, 1994.
- O ponimanii. Opyt issledovaniia prirody, granits i vnutrennego stroeniia nauki kak tsel'nogo znaniia*. Moscow: Tipografiia E. Lissnera i Iu. Romana, 1886. Repr. Moscow: Tanais.
- Poslednie list'ia*, edited by A. N. Nikoliukin. Moscow: Respublika, 2000.
- Religiia i kul'tura*, edited by E. V. Barabanov. 2 vols. Moscow: Pravda, 1990.
- Sakharna*, edited by A. N. Nikoliukin. Moscow: Respublika, 1998.
- Sredi khudozhnikov*, edited by A. N. Nikoliukin. Moscow: Respublika, 1994.
- V temnykh religioznykh luchakh*, edited by A. N. Nikoliukin. Moscow: Respublika, 1994.

### WORKS BY ROZANOV TRANSLATED INTO ENGLISH

- Dostoevsky and the Legend of the Grand Inquisitor*. Translated by Spencer E. Roberts. Ithaca: Cornell University Press, 1972.
- Four Faces of Rozanov: Christianity, Sex, Jews, and the Russian Revolution*. Translated by Spencer E. Roberts. New York: Philosophical Library, 1978.
- Selections from Fallen Leaves*. Translated by Boris Jakim and Laury Magnus. Chicago: The Variable Press, 2003.

### WORKS ABOUT ROZANOV

- Clowes, Edith W. *Fiction's Overcoat: Russian Literary Culture and the Question of Philosophy*. Ithaca and London: Cornell University Press, 2004.
- Crone, Anna Lisa. *Rozanov and the End of Literature: Polyphony and the Dissolution of Genre in Solitaria and Fallen Leaves*. Würzburg: jal-Verlag, 1978.
- Fateev, V. A., ed. *V. V. Rozanov: Pro et contra. Antologiya*. 2 vols. St. Petersburg: Izdatel'stvo Russkogo Khristianskogo gumanitarnogo institute, 1995.
- Hutchings, Stephen C. “Breaking the Circle of the Self: Domestication, Alienation, and the Question of Discourse Type in Rozanov’s Late Writings.” *Slavic Review* 52 (1993): 67–86.
- Lebedev, Iu. V., ed. *Chteniia, posviashchënnnye 80-letiiu pamiati V. V. Rozanova*. Kostroma: Kostromskoi filial Rossiiskogo fonda kul'tury, 1999.
- Nikoliukin, A. N. *Golgofa Vasiliia Rozanova*. Moscow: Russkii put', 1998.
- Rudnev, Pavel. *Teatral'nye vzgliady Vasiliia Rozanova*. Moscow: Agraf, 2003.
- Siniavskii, A. D. *Ocherki russkoi kul'tury. 1. Opavshie list'ia Vasiliia Vasilievicha Rozanova*. Paris: Sintaksis, 1982.

**Robert Bird (2005)**

## RÜDIGER, ANDREAS

(1673–1731)

Andreas Rüdiger, the German physician and philosopher, was born in Rochlitz, Saxony. Poverty and bad health allowed him to study only irregularly. In 1692 he served as a tutor in the home of Christian Thomasius. He was compelled to interrupt his studies completely in 1695; not until 1697 could he enter the University of Leipzig, where he studied law and medicine, receiving a master’s degree in 1700. He received a doctorate in medicine from the University of Halle in 1703, but he continued to lecture at the University of Leipzig. From 1707 to 1712 he practiced medicine and lectured in Halle, and from 1712 until his death he did so in Leipzig.

The development of Rüdiger’s philosophy was greatly influenced by his teachers Christian Thomasius and Franz Budde. However, he soon developed individual views within the Thomasian school. His medical studies centered his interests on natural philosophy and gave his thought a practical bent. Like Budde’s, Rüdiger’s mind was more systematic than Thomasius’s.

Rüdiger’s most important work, *Philosophia Synthetica* (1706–1707), is divided into three sections: “Wisdom,” “Justice,” and “Prudence.” The section on wisdom embraces logic and natural philosophy, that on justice

covers metaphysics and natural law, and that on prudence covers ethics and politics.

Rüdiger's logic had a clear psychological orientation. He was mainly interested in the origin and development of our ideas, which, he held, come into our minds through the senses, although there are some innate mental elements, too. He criticized René Descartes, discussed Pierre Gassendi, and drew some inspiration from John Locke. Rüdiger stressed the passive element of the mind; reflection, or *sensio interna*, is (contrary to Locke) a passive fact. The standard of truth lies in man's consciousness, in a *recta ratio*, which is not common sense but something that can be acquired only through instruction in logic (*lumen acquisitum*). Logic was therefore more important for Rüdiger than for the other members of the Thomasian school. He developed a refined syllogistic theory, formalizing his acceptance of the mathematical method in philosophy. However, he conceived the mathematical method quite differently from Christian Wolff, as a method for deducing facts from given facts rather than as the drawing of possible conclusions from abstract principles. Rüdiger's philosophy, like that of the Thomasian school generally, was based in large part on the notion of reality and appealed mainly to the senses and to experience, both interior and exterior. He defined "truth" in connection with the possibility of perceiving and "existence" in connection with being perceived—again in the tradition of Thomasian subjectivism.

In natural philosophy, Rüdiger tried to combine the Thomasian and Pietistic animistic or spiritualistic physics with mechanism, but the spiritualistic element predominates. He held that we have no certain knowledge of nature, and generally he refrained from choosing between different hypotheses, for instance, between the Copernican and the biblical astronomical theories.

The practical bent of Rüdiger's philosophy explains why he discussed metaphysics under the heading of justice. His metaphysical discussions were largely devoted to theology and to man's duties toward God; his discussions of natural law were devoted to our duties toward other men. Metaphysics is the science of reality, and in particular of the *ens realissimum*, rather than the science of possibility. However, according to Rüdiger, we cannot penetrate the essence of things in metaphysics; we can only establish, by means of experience, that things exist and how they exist.

Rüdiger's section on prudence constitutes, in the Thomasian tradition, a kind of anthropology, both private and public. Ethics provides precepts for reaching

happiness on Earth, and politics provides precepts for governing a commonwealth.

Through his pupil A. F. Hoffmann, Rüdiger exerted a strong influence on the development of the philosophy of Christian August Crusius, and through Crusius on the whole development of German philosophy.

**See also** Crusius, Christian August; Thomasius, Christian.

### Bibliography

#### WORKS BY RÜDIGER

*Disputatio Philosophica de Eo, Quod Omnes Ideae Oriantur a Sensione*. Leipzig, 1704.

*Philosophia Synthetica. Tribus Libris de Sapientia, Justitia et Prudentia. Methodo Mathematicae Aemula Breviter et Succincte in Usam Auditorum Comprehense*. Leipzig, 1707. Reprinted at Halle, 1711, and at Frankfurt, 1717, as *Institutiones Eruditionis* and at Leipzig, 1723, as *Philosophia Pragmatica. Methodo Apodictica et Quoad Eius Licuit Mathematica Conscripta*.

*De Sensu Veri et Falsi*. Halle, 1709.

*Physica Divina Recta Ira ad Ultramque Hominis Felicitatem Naturalem Atque Moralem Ducens*. Frankfurt, 1716.

*Klugheit zu Leben und zu Herrschen*. Leipzig, 1722.

*Christian Wolffens Meinung von dem Wesen der Seele und eines Geistes überhaupt, und Andreas Rüdigers Gegen-Meinung*. Leipzig, 1727. A pamphlet attacking Wolff.

#### WORKS ON RÜDIGER

Carls, W. *Rüdigers Moralphilosophie*. Halle, 1894.

Schepers, Heinrich. *A. Rüdigers Methodologie und ihre Voraussetzungen*. Cologne, 1959.

Wundt, Max. *Die deutsche Schulphilosophie im Zeitalter der Aufklärung*, 82–97. Tübingen: Mohr, 1945.

Giorgio Tonelli (1967)

## RUFUS, RICHARD

(?–after 1259)

Richard Rufus, a thirteenth-century philosopher and theologian, was among the first European medieval authors to study Aristotelian metaphysics, epistemology, and natural philosophy. His lectures on the so-called *libri naturales* date from a period shortly after the effective lapse of the ban on teaching them in 1231 and are among the earliest European commentaries on those works. In 1238, after writing treatises against Averroes and lecturing on Aristotle—at greatest length on the *Metaphysics*—he joined the Franciscan Order, left Paris, and became a theologian.

Rufus's lectures on Peter Lombard's *Sentences* were the first presented by an Oxford bachelor of theology. Greatly influenced by Robert Grosseteste, Rufus's Oxford lectures were devoted in part to a refutation of Richard Fishacre, the Dominican master who first lectured on the *Sentences* at Oxford.

Rufus's Oxford lectures were employed as a source by St. Bonaventure, whose lectures on the *Sentences* were vastly influential. Returning to Paris shortly after Bonaventure lectured there, Rufus took Bonaventure's lectures as a model for his own Parisian *Sentences* commentary. Rufus's Paris lectures made him famous. According to his enemy Roger Bacon, when he returned to Oxford after 1256 as the Franciscan regent master, his influence increased steadily. It was at its height forty years later in the 1290s, when John Duns Scotus was a bachelor of theology. Early versions of many important positions developed by Duns Scotus can be found in Rufus's works.

## NATURAL PHILOSOPHY

Among the first medieval European philosophers to encounter Aristotle's arguments for the eternity of the world, Rufus also presented some of the most cogent counterarguments. One argument is based on a contradiction between the definitions of *past* and *infinity*. It is impossible to traverse an infinity, but it pertains to the nature of the past to have been traversed; therefore, past time cannot be infinite. In 1235 Rufus presents the argument with characteristic brevity: "Having been traversed" is incompatible with the definition of infinity, but "having been traversed" belongs to the definition of the past. Therefore, being past is incompatible with the definition of infinity.

This argument, first presented in late antiquity by John Philoponus, is now associated with Immanuel Kant; in medieval philosophy it is ordinarily ascribed to Bonaventure, who advanced it in 1250 or 1251. It occurs in different versions, some more persuasive than others. Grosseteste, for example, mistakenly seeks to apply it to the future as well as to the past, claiming that the argument can be used to show that time could not be infinite *a parte post*. Rufus sees even more clearly than Philoponus that the direction of time is an important part of this argument. He notices that the argument must be based on the fact that the whole of past time has been traversed, rather than on the claim that the whole of the past and the future will have been traversed. In his later work he seeks to force his opponents to see that they are committed to the claim that some past days are not now and never were present. By contrast, Philoponus sees this as an argument about the impossibility of completely

counting an infinite series, with no particular focus on the direction of time.

Rufus's version of another of Philoponus's arguments is based on the concept of priority. If the number of days before today is infinite, and the number of days before tomorrow is infinite, then the number of days before today is not less than the number of days before tomorrow. Consequently, today does not arrive sooner than tomorrow, which is absurd. Rufus assumes here that unequal infinities are impossible. Following Georg Cantor, modern mathematicians reject this assumption. However, Rufus needs only the uncontroversial claim that mappable infinities are equal: If one postulates beginningless time, the number of days before today and the number of days before tomorrow are mappable infinite series. Rufus might still argue that if the world has no beginning, then one must give up the belief that less time elapses before earlier events than before later events.

Philoponus's original version of this argument is not based on the claim that more time transpires before later events than before earlier events. The absurdities he asks one to reject are mathematical: that it is possible to add to an infinity, or that one infinity can be multiplied by another, so that one infinity would be greater than another by a determinate proportion. By contrast, the absurd conclusion Rufus asks one to reject is that "today does not come sooner than tomorrow"; he emphasizes the unique properties of time.

## THEORY OF KNOWLEDGE

The fullest statement of Rufus's epistemological views now known is a treatise titled *Speculum animae* (A mirror of the soul), probably written to explain problems in Aristotelian philosophy to his Franciscan confreres. This treatise addresses the question: What does Aristotle mean when he says that "in some manner the soul is every thing"? In the *Speculum* Rufus develops and changes his views; he rejects the view his predecessors based on patristic authorities: The soul is everything because it shares being with rocks, life with animals, and understanding with angels—a view Rufus states without comment in the last lectures he gave before becoming a Franciscan, when expounding *Metaphysics* Lambda. Since Rufus also rejects a literal interpretation of the dictum, he must explain in what sense the soul becomes an object when it understands or senses that object.

Rufus has to face two related questions: Why does the soul not become green when it perceives something green? If reception of species produces apprehension in the soul, why does the presence of such species not have

this effect in other subjects? One element of Rufus's reply is constant. He postulates that sensible and intelligible species are nonnatural and different in their mode of being from external objects. Such species are not described by Aristotelian categories; they are neither substances nor accidents. Accordingly, their reception does not produce the object sensed but the sensation or cognition of the relevant object; the sensitive soul does not become green; it senses the color green. In his *De anima* commentary Rufus describes the direct objects of sensation as spiritual beings, and he holds that plants do not sense colors since spiritual beings do not act on them. Spiritual being, a concept Rufus owes to Averroes, is the key to Rufus's exposition of a phrase in Aristotle, who says that the senses are susceptible of sensible species "without matter" (2.12.424a18–19). In his *Contra Averroem* Rufus confronts the objection that accidents that are not spiritual also act without matter—in producing heat, for example. He replies by claiming that Aristotle was contrasting species or intentions with "materiated species" designed to perfect matter rather than to produce cognition. In his last *Metaphysics* commentary Rufus contrasts spiritual with material reception, repeating terminology from his *In De anima*, but omitting the *De anima* commentary's reference to spiritual being.

In his *Contra Averroem* Rufus also makes it harder to answer the second question in another respect. Lecturing on *De anima*, Rufus claims that wood, for example, apprehends nothing because its matter receives only the natural form of wood, not its species (a similitude of the whole). In *Contra Averroem* he cites passages that convince him that the objects one senses are not mere similitudes of sensed objects but really the same as them. Rufus's response to the problem this presents in the *Speculum animae* (and subsequently in his Oxford theology lectures) is to argue that what is really identical may be formally distinct. Since species exist nonnaturally, they can be really the same as, but not formally identical with, or predicable of, the objects of apprehension. This safeguards the claim that what one apprehends is really the same as external objects; in some sense the soul really is all things.

## METAPHYSICS

Postulating a kind of identity that permits real but not formal predication is a conceptual tool that Rufus employs when discussing a variety of philosophical topics—for example, the problem of individuation. Like Duns Scotus in his *Metaphysics* commentary, Rufus postulates individual forms to explain individuation. Indi-

vidual forms are really, but not formally, the same as specific forms. Specific forms are principles of shared identity; they pertain to common natures capable of instantiation (*multiplicabilis*). By contrast, individual forms pertain to the same natures as they are actually instantiated (*actu multiplicata*).

Rufus's arguments against alternative theories were initially more influential than his own views. He holds that the cause of individuation cannot be an accident or an aggregation of accidents, since individual primary substances are ontologically prior to accidents. Though he allows a role for matter as an occasional cause of individuation, he argues that even determinate matter could not by itself be the principle of individuation. Being an individual means being distinct and united, both of which are functions of form, the active principle of substance, not matter, the passive principle.

Holding that individual forms added to an aggregate of matter and specific form must be the principle of individuation, Rufus denies that the ultimate constituents of individuals are knowable. He is not sure whether what is added to the common nature can be located within an Aristotelian category. He suggests that, strictly speaking, the cause of individuation may be neither a substance nor an accident. Identifying individual forms as perfections of the specific form, he suggests that they may be substantial without being substances. Specific and individual forms provide different degrees of unity: Specific unity is less than individual unity and greater than generic unity.

Like Rufus's views on individuation, his argument for the existence of God was accepted and modified by Duns Scotus. Rufus rejected St. Anselm's famous ontological argument as sophistical (though subtle). In its place he advanced a modal argument based on the concept of God as an independent being (*a se et non ab alio*). The existence of independent beings is either necessary or impossible. Therefore, if an independent being can exist, it does exist. Rufus employs logically sophisticated arguments to show that an independent being can exist.

At the opposite end of the cosmological scale, Rufus also makes an original contribution to the problem of elemental composition—positing elemental forms incompletely actualized in the compounds they comprise. Rufus sets out to explain how elements can retain their identity in a compound, so that we can correctly say that a compound is composed of elements, and yet also explain the unity of the compound with a distinct identity. Rufus has to describe how flesh and bone can be composed of elements that can be separated out when the compound breaks down, without immediately being dis-

solved by their component elements' actions on one another—the hot heating the cold, for example. If there is to be a compound at all, the elements cannot exist in the compound in quite the same way that they do when separated. Supposing that the elements are substances, Rufus argues contrary to Averroes that elemental forms are in no sense accidents, either in the compound or outside it, though they are subject to intension and remission. His solution to the difficulty is to postulate that the elemental forms can be more or less actual; they exist in compounds in accidental or proximate potential, prevented from complete actuality by the presence of the contrary elements. The resulting mixture Rufus describes as having the unity of fusion, intermediate between absolute unity and unity of aggregation.

## INFLUENCE

Rufus's importance has long gone unrecognized, in part because he preferred not to take credit for his own work and in part because, unlike his contemporaries, he provided long quotations of the positions he treated seriously. Since his own views were often stated briefly, historians who overlooked his critical bent saw him as a derivative figure. Now that Bonaventure's borrowing from Rufus has been discovered, and scholars are beginning to appreciate the significance of citations by Grosseteste (to Magister Richardus), Duns Scotus (to *Doctor antiquus*), and Franciscus de Marchia (to Richardus), the question of Rufus's influence will have to be reconsidered.

**See also** Anselm, St.; Bacon, Roger; Bonaventure, St.; Duns Scotus, John; Eternity; Grosseteste, Robert; Philoponus, John.

## Bibliography

### WORKS BY RUFUS

*In Physicam Aristotelis*, edited by Rega Wood. New York: Oxford University Press, 2003.

[Pseudo-Peter of Spain]. *Expositio libri De anima*. In Pedro Hispano, *Obras filosóficas*. Vol. 3, edited by M. Alonso. Instituto de Filosofía "Luis Vives," Serie A, no. 4. Madrid: n.p., 1952.

### WORKS ABOUT RUFUS

Brams, J. "Le premier commentaire médiéval sur le 'Traité de l'âme' de Aristote?" *Recherches de théologie et philosophie médiévales* 68 (2001): 213–227.

Brown, S. F. "The Eternity of the World Discussion at Early Oxford." In *Mensch und Natur im Mittelalter*. Vol. 1, edited by Albert Zimmermann and Andreas Speer, 1: 259–280. Berlin and New York: De Gruyter, 1991.

Dales, R. C. "The Influence of Grosseteste's Hexaëmeron on the Sentences Commentaries of Richard Fishacre O.P. and

Richard Rufus of Cornwall O.F.M." *Viator* 2 (1971): 271–300.

Gál, G. "Commentarius in 'Metaphysicam' Aristotelis cod. Vat. lat. 4538 fons doctrinae Richardi Rufi." *Archivum franciscanum historicum* 43 (1950): 209–242.

Gál, G. "Opiniones Richardi Rufi Cornubiensis a censore reprobatae." *Franciscan Studies* 35 (1975): 136–193.

Gál, G. "Viae ad existentiam Dei probandam in doctrina Richardi Rufi OFM." *Franziskanische Studien* 38 (1956): 177–202.

Karger, E. "Richard Rufus on Naming Substances." *Medieval Philosophy and Theology* 7 (1998): 51–67.

Karger, E. "Richard Rufus's Account of Substantial Transmutation." *Medioevo* 27 (2002): 165–189.

Libera, A. de. "La littérature des abstractions et la tradition logique d'Oxford." In *The Rise of British Logic: Acts of the Sixth European Symposium on Medieval Logic and Semantics, Balliol College, Oxford, 19–24 June 1983*, ed. P. Osmund Lewry, 63–114. Papers in Mediaeval Studies, no. 7. Toronto, Canada: Pontifical Institute of Mediaeval Studies, 1985.

Little, A. G. "The Franciscan School at Oxford in the Thirteenth Century." *Archivum franciscanum historicum* 19 (1926): 803–874.

Long, R. J. "The First Oxford Debate on the Eternity of the World." *Recherches de philosophie et théologie médiévales* 65 (1998): 54–98.

Long, R. J. "Of Angels and Pinheads: The Contributions of the Early Oxford Masters to the Doctrine of Spiritual Matter." *Franciscan Studies* 56 (1998): 239–254.

Noone, T. B. "An Edition and Study of the *Scriptum super Metaphysicam*, Bk. 12, Dist. 2: A Work Attributed to Richard Rufus of Cornwall." PhD diss., University of Toronto, 1987.

Noone, T. B. "Richard Rufus of Cornwall and the Authorship of the *Scriptum super Metaphysicam*." *Franciscan Studies* 49 (1989): 55–91.

Noone, T. B. "Richard Rufus on Creation, Divine Immutability, and Future Contingency in the *Scriptum super Metaphysicam*." *Documenti e studi sulla tradizione filosofica medievale* 4 (1993): 1–23.

Noone, T. B. "Roger Bacon and Richard Rufus on Aristotle's *Metaphysics*: A Search for the Grounds of Disagreement." *Vivarium* 35 (1997): 251–265.

Noone, T. B., and R. J. Long. "Fishacre and Rufus on the *Metaphysics of Light*: Two Unedited Texts." In *Roma, magistra mundi: Itineraria culturae mediaevalis: Mélanges offerts au père L. E. Boyle à l'occasion de son 75e anniversaire*, 3 Vols., edited by Jacqueline Hamesse, 2:517–548. Textes et études du moyen âge, no. 10. Louvain-la-Neuve, Belgium: Fédération des Instituts d'Etudes Médiévales, 1998.

Pelster, F. "Der älteste Sentenzenkommentar aus der Oxforder Franziskanerschule: Ein Beitrag zur Geschichte des theologischen Lehrbetriebs an der Oxforder Universität." *Scholastik* 1 (1926): 50–80.

Pelster, F. "Richardus Rufus Anglicus O.F.M. (c. 1250), ein Vorläufer des Duns Scotus in der Lehre von der Wirkung der priesterlichen Lossprechung." *Scholastik* 25 (1950): 549–552.

Plevano, R. "Richard Rufus of Cornwall and Geoffrey of Aspill: Two Questions on the Instant of Change." *Medioevo* 19 (1993): 167–232.

- Raedts, P. *Richard Rufus of Cornwall and the Tradition of Oxford Theology*. Oxford Historical Monographs. Oxford, 1987.
- Weisberg, M., and R. Wood. "Richard Rufus's Theory of Mixture: A Medieval Explanation of Chemical Combination." In *Chemical Explanation: Characteristics, Development, Autonomy*, edited by J. E. Earley, Sr., 282–292. Annals of the New York Academy of Sciences, vol. 988. New York, 2003.
- Wood, R. "Distinct Ideas and Perfect Solicitude: Alexander of Hales, Richard Rufus, and Odo Rigaldus." *Franciscan Studies* 53 (1993): 7–46.
- Wood, R. "The Earliest Known Surviving Western Medieval Metaphysics Commentary." *Medieval Philosophy and Theology* 7 (1998): 39–49.
- Wood, R. "Early Oxford Theology." In *Mediaeval Commentaries on the Sentences of Peter Lombard: Current Research*, edited by G. R. Evans, 1: 289–343. Leiden: E.J. Brill, 2002.
- Wood, R. "Individual Forms: Richard Rufus and John Duns Scotus." In *John Duns Scotus: Metaphysics and Ethics*, edited by L. Honnefelder, R. Wood, and M. Dreyer, 251–272. Leiden: E.J. Brill, 1996.
- Wood, R. "Richard Rufus: Physics at Paris before 1240." *Documenti e studi sulla tradizione filosofica medievale* 5 (1994): 87–127.
- Wood, R. "Richard Rufus and the Classical Tradition: A Medieval Defense of Plato." In *Néoplatonisme et philosophie médiévale: Actes du Colloque international de Corfou 6–8 octobre 1995 organisé par la Société Internationale pour l'Étude de la Philosophie Médiévale*, edited by L. G. Benakis, 229–251. Rencontres de philosophie médiévale, no. 6. Turnhout, 1997.
- Wood, R. "Richard Rufus of Cornwall and Aristotle's Physics." *Franciscan Studies* 52 (1992): 247–281.
- Wood, R. "Richard Rufus of Cornwall on Creation: The Reception of Aristotelian Physics in the West." *Medieval Philosophy and Theology* 2 (1992): 1–30.
- Wood, R. "Richard Rufus's De anima Commentary: The Earliest Known, Surviving, Western De anima Commentary." *Medieval Philosophy and Theology* 10 (2001): 119–156.
- Wood, R. "Richard Rufus' 'Speculum animae': Epistemology and the Introduction of Aristotle in the West." In *Die Bibliotheca Amploniana: Ihre Bedeutung im Spannungsfeld von Aristotelismus, Nominalismus und Humanismus*, edited by A. Speer, 86–109. Miscellanea mediaevalia, vol. 23. Berlin, 1995.
- Wood, R. "Roger Bacon: Richard Rufus' Successor as a Parisian Physics Professor." *Vivarium* 35 (1997): 222–250.
- Wood, R., and R. Andrews. "Causality and Demonstration: An Early Scholastic Posterior Analytics Commentary." *The Monist* 79 (1996): 325–356.
- Wood, R., and M. Weisberg. "Interpreting Aristotle on Mixture: Problems about Elemental Composition from Philoponus to Cooper." *Studies in History and Philosophy of Science* 35 (2004): 681–706.

**Rega Wood and Jennifer R. Ottman (2005)**

## RULE FOLLOWING

In 1982 Saul Kripke published *Wittgenstein on Rules and Private Language* and ushered in a new era of Ludwig Wittgenstein interpretation. Although elements of Kripke's view of Wittgenstein could be found in the preceding literature (notably in Robert Fogelin's *Wittgenstein*), nothing had captured attention like his presentation of the "rule-following considerations."

Kripke presented his essay as a reconstruction of the problems Wittgenstein is addressing between around §140 and §203 of the *Philosophical Investigations*. These issue in the form of a paradox—that there can be no such thing as the meaning of a word; no fact of the matter that entails that a word is used according to a rule, whereby some applications of it are determined to be correct and other applications incorrect. In §201 Wittgenstein wrote "This [is] our paradox: no course of action could be determined by a rule, because every course of action can be made out to accord with the rule. The answer [is] if everything can be made out to accord with the rule, then it can also be made out to conflict with it. And so there would be neither accord nor conflict here."

The paradox is developed by Kripke through the figure of a "bizarre skeptic." The defender of common sense, here the view that words do indeed have meanings and obey rules, is challenged to show what this meaning consists in. The facts he or she can adduce typically include past applications and present dispositions to apply words in new cases. They may also include flashes of consciousness—for instance, if we associate a particular image with a term. But, Kripke's skeptic argues, these are not the kinds of facts that can determine the actual rule that governs the meaning of a word. The skeptic adduces three kinds of problems. First, our dispositions are finite, whereas a rule can cover a potential infinity of new cases. Second, our dispositions sometimes fail to match the relevant rules: This is precisely what happens when we mistakenly apply words to things to which they do not in fact apply. Third, the existence of a rule has normative implications. It determines correctness and incorrectness of application of the term it governs. Our dispositions, by contrast, have no such implication. There is nothing intrinsically wrong about bending our dispositions from moment to moment, in the way that there is about applying a term in a way that fails to accord with its meaning. Finally, the addition of flashes of consciousness is unlikely to help, for, as Wittgenstein himself said, any such fact itself stands in need of interpretation. A flash of consciousness cannot comprehend all the possible appli-

cations of a term and sort them into those that are correct and those that are not.

Kripke illustrates these points with the case of a strange arithmetical operator, “quus.” For two numbers  $n$  and  $m$ ,  $n$  quus  $m$  is identical with  $n$  plus  $m$  for sufficiently small or common numbers, but the two results (or calculations) diverge when  $n$  and  $m$  are greater than a certain value (the function is therefore reminiscent of Nelson Goodman’s predicate, “grue”). We do not mean  $n$  quus  $m$  when we talk of  $n$  plus  $m$ . But our dispositions with “plus” might match those of people who in fact use the term to mean quus; we might give the answer  $n$  quus  $m$  when we attempt to add  $n$  and  $m$ , since we make mistakes; and finally there is nothing right or wrong about having one disposition or another.

The conclusion is paradoxical, since nothing seems more certain than that we do succeed in attaching reasonably determinate meanings to terms. It may be true that the “open texture” of terms suggests that meanings are never fully precise, capable of determining their application in any circumstances, however outlandish. Nevertheless, over an indefinite normal range of cases, there is no doubt that some applications are correct and others not, and any interpretation of us according to which we mean something along the lines of the “quus” function is incorrect. Yet so long as the skeptic wins, we have no conception of our right to say such things. Kripke’s own solution to the paradox is that the skeptic wins on his chosen ground. There is indeed no fact of the matter whether one rule rather than another governs the use of a term. But we can advance a “skeptical solution” (David Hume’s phrase from a different context) to the doubts. What there is instead is a practice of regarding ourselves and others in certain lights. We dignify each other as meaning one thing or another by our terms, and this ongoing practice is all that there is.

Kripke’s work generated enormous interest and a variety of responses in the literature. Some outraged students of Wittgenstein argued that it was not at all his intention to produce a paradox but to lay bare the oversimplifications, or desire for a simple theory, that trap people into finding rule following problematic (Baker and Hacker 1984). Many writers queried whether Wittgenstein could consistently have been content with a “non-truth-conditional” account of rule following, which is what Kripke offers him, since Wittgenstein’s abhorrence of theory and his belief that philosophy leaves everything as it is would make it impossible for him to say that it is not strictly speaking true that the application of words is correct or incorrect. Some (McDowell 1981)

detected a mischievous dislike of soft, humanly oriented facts in the setting up of the paradox and argued that a proper appreciation of the human constitution of rule following had wide implications for the notion of objectivity, as it occurs in domains such as aesthetics or ethics. Some (McGinn 1984) found that Kripke had not looked hard enough for natural facts with which to identify the obtaining of a rule; others (Blackburn 1985) embraced the thought that since the loss of a normative element in meaning was the main problem underlying the paradox, and since naturalistic theories of normativity have been proposed in many guises, a more generous sense of how to talk about facts solves the paradox. Paul Boghossian (1989) provided a summary of the state of the debate and a controversial contribution to it.

**See also** Goodman, Nelson; Hume, David; Kripke, Saul; Philosophy of Language; Wittgenstein, Ludwig Josef Johann.

### **Bibliography**

- Baker, G. P., and P. M. S. Hacker. *Scepticism, Rules, and Language*. Oxford: Blackwell, 1984.
- Blackburn, S. “The Individual Strikes Back.” *Synthese* 20 (1985).
- Boghossian, P. “The Rule-Following Considerations.” *Mind* 98 (1989).
- McDowell, J. “Non-Cognitivism and Rule-Following.” In *Wittgenstein: To Follow a Rule*, edited by S. Holtzman and C. Leich. London: Routledge and Kegan Paul, 1981.
- McGinn, C. *Wittgenstein on Meaning*. Oxford: Blackwell, 1984.
- Wittgenstein, L. *Philosophical Investigations*. Oxford: Blackwell, 1953.

**Simon Blackburn (1996)**

## RULE FOLLOWING [ADDENDUM]

Paul Boghossian (1990) summarizes four components of what has become the received interpretation of Saul Kripke’s book *Wittgenstein on Rules and Private Language* (1982) (*WRPL*): (i) Kripke’s meaning-skeptic argues that meaning is normative and from descriptive facts one cannot derive normative claims, so descriptive facts cannot explain or reduce meaning; (ii) nonreductive accounts of meaning are not vulnerable to the meaning-skeptic’s argument; (iii) Kripke’s skeptical solution is a kind of nonfactualism about meaning; and (iv) Kripke’s arguments do not show that an isolated individual cannot follow rules. Contributions to the literature about



rule-following published after 1989 include criticisms of the meaning-skeptic's argument, investigations of the nature and coherence of nonfactualism about meaning, revisionary interpretations of *WRPL*, and interpretations of Wittgenstein that challenge Kripke's and others' understanding of rule-following.

Ruth Garrett Millikan (1990) argues that the normativity of meaning can be explained in terms of biological purposes shaped and sustained by natural selection. Philip Pettit (1990) argues that the normativity of meaning is explained by one's dispositions to question or revise a judgment when one has reason to think it was the result of interfering factors, such as poor lighting or intoxication. According to Paul Horwich (1995), there is a sense of "determine" in which facts about the use of a linguistic expression *E* can determine the meaning of *E*, even if the meaning of *E* cannot be "read-off" from those facts, as Kripke's meaning-skeptic argues.

Boghossian (1990) argues that nonfactualism about meaning is incoherent because it simultaneously *presupposes* and *conflicts with* deflationism about truth. Robert Kraut (1993) responds to this argument by describing a kind of deflationism about truth that seems compatible with nonfactualism about meaning. Crispin Wright (1992) and Scott Soames (1999) also identify what they regard as errors in Boghossian's argument.

George Wilson (1992) criticizes (iii), citing passages of *WRPL* that suggest that ascriptions of meaning can be factual even if they are not determined by independently specifiable facts. Against (i), José Zalabardo (1997) argues that the heart of the meaning-skeptic's argument is not that from descriptive facts one cannot derive normative claims, but that no facts of which we can be immediately aware can determine that our application of a word to a new case is justified. Gary Ebbs (1997) criticizes both (i) and (ii) by reconstructing Kripke's skeptical reasoning in a way that does not presuppose reductionism about meaning. Donald Davidson (1992) qualifies (iv) by supplementing Kripke's account of meaning-ascriptions with the premise that meaning requires triangulation.

Donna Summerfield (1990) reads Wittgenstein as presenting a positive account of representation that does not apply to an individual in isolation. Edward Minar (1991) presents an alternative to Kripke's skeptical solution and to the communitarian view that rule-following is constituted by linguistic interactions between speakers. David Bloor (1997) defends a communitarian interpretation of Wittgenstein. David Finkelstein (2000) suggests that to understand Wittgenstein one must come to see

that there is not always a gulf between a rule and its application.

**See also** Davidson, Donald; Kripke, Saul; Meaning; Millikan, Ruth; Private Language Problem; Wittgenstein, Ludwig Josef Johann.

### **Bibliography**

- Bloor, D. *Wittgenstein: Rules and Institutions*. London: Routledge, 1997.
- Boghossian, P. "The Status of Content." *The Philosophical Review* 99 (1990).
- Davidson, D. "The Second Person." *Midwest Studies in Philosophy* 17 (1992).
- Ebbs, G. *Rule-Following and Realism*. Cambridge, MA.: Harvard University Press, 1997.
- Finkelstein, D. "Wittgenstein on Rules and Platonism." In *The New Wittgenstein*, edited by A. Crary and R. Read. London: Routledge, 2000.
- Goldfarb, W. "Kripke on Wittgenstein on Rules." *Journal of Philosophy* 83 (1985).
- Horwich, P. "Meaning, Use, and Truth." *Mind* 104 (1995).
- Kraut, R. "Robust Deflationism." *Philosophical Review* 102 (1993).
- Kripke, S. *Wittgenstein on Rules and Private Language*. Cambridge, MA.: Harvard University Press, 1982.
- McDowell, J. "Wittgenstein on Following a Rule." *Synthese* 58 (1984).
- Miller, A., and C. Wright. *Rule-Following and Meaning*. Montreal: McGill-Queen's University Press, 2002.
- Millikan, R. "Truth, Hoverflies, and the Kripke-Wittgenstein Paradox." *Philosophical Review* 99 (1990).
- Minar, E. "Wittgenstein and the 'Contingency' of Community." *Pacific Philosophical Quarterly* 72 (1991).
- Pettit, P. "The Reality of Rule-Following." *Mind* 99 (1990).
- Soames, S. *Understanding Truth*. New York: Oxford University Press, 1999.
- Summerfield, D. "Philosophical Investigations 201: A Wittgensteinian Reply to Kripke." *Journal of the History of Philosophy* 28 (1990).
- Wilson, G. "Kripke on Wittgenstein on Normativity." *Midwest Studies in Philosophy* 17 (1992).
- Wright, C. *Truth and Objectivity*. Cambridge, Mass.: Harvard University Press, 1992.
- Zalabardo, J. "Kripke's Normativity Argument." *Canadian Journal of Philosophy* 27 (1997).

**Gary Ebbs (2005)**

## RULE-UTILITARIANISM

See *Utilitarianism [addendum]*

## RUSKIN, JOHN

(1819–1900)

John Ruskin, the English critic of art and society, was born in London, the son of a wine merchant. He began writing while at Oxford and in 1843 published, in London, the first volume of *Modern Painters*, four more volumes of which were published during the next sixteen years. In 1849 he published *The Seven Lamps of Architecture* and between 1851 and 1853 *The Stones of Venice* (3 vols.). The major part of his work as a young man was criticism of art and architecture, and his subsequent ethical and social writing grew from this root. The beginnings of this important extension of his range can be seen in the famous chapter “The Nature of Gothic” in *The Stones of Venice*; the important connection established there, between art and “the right kind of labour,” is developed in *The Political Economy of Art* (printed as *A Joy for Ever*, 1857), *Unto This Last* (1862), *The Crown of Wild Olive* (1866), and *Munera Pulveris* (1863 and 1872). Meanwhile Ruskin continued his criticism of art and architecture, notably in *The Two Paths* (1859) and in his lectures as Slade professor of art at Oxford, between 1870 and 1879 and in 1883/1884. A volume of essays, *Sesame and Lilies*, appeared in 1865 and an unfinished autobiography, *Praeterita*, between 1885 and 1889. He also published letters on social questions, notably in *Time and Tide* (1867) and *Fors Clavigera* (8 vols., 1871–1884).

Ruskin’s social and ethical teaching, though deeply influenced by the work of Thomas Carlyle, followed from his understanding of the nature of art. The artist’s function is to reveal aspects of the universal truth, which is also beauty. Any corruption of the moral nature of the artist is an inevitable corruption of this revelation, but it is impossible, finally, for an artist to be good if his society is corrupt. The art of any society is, correspondingly, “the exact exponent of its social and political virtues.” Where there is a lack of “wholeness” in art (wholeness being a full and deep response to the organic life of the universe), there is a corresponding lack of “wholeness” in society; to recover the one men must recover the other. Just as the beauty of art is the expression of the essential nature of the universe—what Ruskin called “typical beauty”—so the goodness of man is the “exertion of perfect life,” which, in comparable relation to the grand design of the created universe, is no more and no less than the “felicitous fulfillment of function” in all living things.

From his work on Venice, Ruskin developed a comparative historical approach to the social conditions in which the “exertion of perfect life” can be fostered or

damaged. In particular, following the English romantic writers and the architectural critic A. W. Pugin, he saw nineteenth-century industrial civilization as the enemy of wholeness in its rampant individualism, its substitution of “production” for “wealth,” and its basic misunderstanding of the nature of work. This kind of social criticism came in many respects to resemble the ideas of some philosophical socialists, and Ruskin’s work had an important formative influence on the British labor movement, both directly and through his influence on William Morris, who united Ruskin’s ideas with a direct commitment to socialism.

Ruskin’s opposition to individualism as a social principle and to competition as a method of political economy was based on his idea of function, the fulfillment of each man’s part in the general design of creation. This required a social order based on intrinsic human values, whereas the existing social order, based on the supposed laws of supply and demand, tended to put the economy above men—indeed, to reduce them to mere “labor”—and, by separating work from the pursuit of human perfection, to separate the work from the man, producing only an alienated and fragmented being. Wherever value is understood as “exchange value,” rather than as the “intrinsic value” derived from function in the universal design, this corruption of man to a mere tool or machine is inevitable. In particular, the confusion about the nature of value leads to false definitions of both wealth and labor. Labor is degraded whenever it is anything other than the “exertion of perfect life,” a creative activity comparable to that of the artist. Wealth is degraded whenever it is confused with mere production, for the meaning of wealth is human well-being, which in material terms is “the possession of useful articles *which we can use*.” Even if the existing system always produced useful articles, the kind of society that it also produced made just distribution and wise consumption impossible. Much actual production, and its widespread misuse, could more properly be called “illth” than “wealth,” for if it possessed only exchange value and not intrinsic value it corrupted its makers and its users.

The most remarkable aspect of Ruskin’s work, then, is the development of a philosophy of art into a moral critique of industrial capitalism. It is a very individual achievement, but it is also part of a general movement of nineteenth-century English thought and has evident connections with William Wordsworth, Percy Bysshe Shelley, Samuel Taylor Coleridge, and Carlyle, as well as with Morris and the Guild Socialists whom Ruskin so notably influenced.

*See also* Aesthetics, History of; Beauty; Carlyle, Thomas; Coleridge, Samuel Taylor; Shelley, Percy Bysshe; Socialism.

### **Bibliography**

#### WORKS BY RUSKIN

*Works*. Edited by E. T. Cook and Alexander Wedderburn, 39 vols. London: Allen, 1903–1912.

#### WORKS ON RUSKIN

- Evans, Joan. *John Ruskin*. London: Cape, 1954.  
 Hobson, J. A. *John Ruskin, Social Reformer*. Boston: Estes, 1898.  
 Hough, Graham. *The Last Romantics*. London: Duckworth, 1949.  
 Quennel, Peter. *John Ruskin, Portrait of a Prophet*. London, 1949.  
 Whitehouse, J. H. *Ruskin the Prophet*. London: Allen and Unwin, 1920.  
 Williams, Raymond. *Culture and Society*, Part 1, Ch. 7. New York: Columbia University Press, 1958.

*Raymond Williams (1967)*

## RUSSELL, BERTRAND ARTHUR WILLIAM (1872–1970)

Bertrand Arthur William Russell, the British philosopher, mathematician, and social reformer, was born in Trelleck, Wales. He was the grandson of Lord John Russell, who introduced the Reform Bill of 1832 and later twice served as prime minister under Queen Victoria. John Stuart Mill, a close friend of Russell's parents, was his godfather in an informal sense. Russell's parents died when he was a little child. Both of them had been freethinkers, and his father's will had provided that he and his brother were to have as their guardians friends of his father's who shared the latter's unorthodox opinions. As the result of litigation the will was set aside by the Court of Chancery and the two boys were placed in the care of their paternal grandparents. Lord John Russell died two years later, and it was the boys' grandmother who determined the manner of their upbringing. Russell was not sent to school but received his early education from a number of Swiss and German governesses and, finally, English tutors. He entered Cambridge University in October 1890 and studied mathematics and philosophy at Trinity College from 1890 to 1894. He was a fellow of Trinity College from 1895 to 1901 and lecturer in philosophy there from 1910 to 1916. In 1916 Russell was dismissed by Trinity College

because of his pacifist activities. He was reinstated in 1919 but resigned before taking up his duties.

What is generally considered Russell's most important work in philosophy was done between 1900 and the outbreak of the first world war. From 1916 until the late 1930s Russell did not hold any academic position and supported himself by writing and public lecturing. During this period he wrote some of his most influential books on social questions, including *Marriage and Morals* (London, 1929) and his two books on education—*On Education, Especially in Early Childhood* (London, 1926) and *Education and the Social Order* (London, 1932). These views were put into practice in Russell's experimental school, the Beacon Hill School, which he started with his second wife, Dora, in 1927. Russell left the school in 1934 after he and Dora were divorced (the school itself continued until 1943). Russell returned to more concentrated work in philosophy around 1936. He moved to the United States in 1938, teaching first at the University of Chicago and then at the University of California at Los Angeles. In 1940 he accepted an invitation from the Board of Higher Education of New York City to join the department of philosophy at City College. However, he never had an opportunity to take up this appointment, having been found unfit for this position in a remarkable opinion by a judge who felt he had to protect "public health, safety and morals." From 1941 until 1943 Russell lectured at the Barnes Foundation in Philadelphia (these lectures were later expanded into *A History of Western Philosophy*). Dr. Albert Barnes, the head of this foundation, dismissed Russell in January 1943, on three days' notice. In this instance Russell successfully brought action for wrongful dismissal. In 1944 he returned to Cambridge where he had been reelected to a fellowship at Trinity College.

Russell was a candidate for Parliament on three occasions and was defeated each time: In 1907 he ran at Wimbledon as a candidate of the National Union of Women's Suffrage Societies, in 1922 and 1923 he stood as the Labour Party candidate for Chelsea. Russell was twice jailed—in 1918 for six months on a count of an allegedly libelous article in a pacifist journal and in 1961, at the age of eighty-nine, for one week, in connection with his campaign for nuclear disarmament.

In 1908 Russell was elected a fellow of the Royal Society. He became an honorary fellow of the British Academy in 1949, and in the same year he was awarded the Order of Merit. Russell twice served as president of the Aristotelian Society and was for many years president of the Rationalist Press Association. In 1950 he received the

Nobel Prize for literature. In making the award, the committee described him as “one of our time’s most brilliant spokesmen of rationality and humanity, and a fearless champion of free speech and free thought in the West.”

Russell had three children and was married four times. In 1931, upon the death of his brother, he became the third earl Russell.

Writing in 1935 the German historian Rudolf Metz referred to Russell as “the only British thinker of the age who enjoys world-wide repute.” At that time his works could not circulate in Germany, Italy, or Russia. Now they are available in every major and a great number of minor languages (a truncated version of *A History of Western Philosophy* was allowed to circulate even in the Soviet Union). It is safe to say that not since Voltaire has there been a philosopher with such an enormous audience. Russell also shares with Voltaire a glittering and graceful prose style and a delicious sense of humor. It is perhaps Russell’s humorous irreverence as much as the substance of his heretical opinions that has so deeply offended several generations of moralists and religious conservatives.

In the following section we shall briefly recount some of the highlights and formative influences in Russell’s eventful life and sketch his views on political and social issues. Although these views are certainly logically independent of his more technical work as a philosopher, they deal with questions that have traditionally been discussed by philosophers, and they also help one to understand the basic motives inspiring Russell’s thought.

## LIFE AND SOCIAL THEORIES

Russell’s childhood and adolescence were unhappy. The atmosphere in his grandmother’s house was one of puritan piety and austerity, and his loneliness, he recalls, was almost unbearable. Only virtue was prized—“virtue at the expense of intellect, health, happiness, and every mundane good.” At the age of five Russell reflected that if he lived to be seventy, he had endured only a fourteenth part of his life, and he felt the long-spread-out boredom ahead of him to be unendurable. In adolescence, he remarks, he was continually on the verge of suicide, from which, however, he was “restrained by the desire to know more mathematics.” His grandmother had gradually moved from Scottish Presbyterianism to Unitarianism. As a child Russell was taken on alternate Sundays to the parish church and to the Presbyterian Church, while at home he was taught the tenets of Unitarianism. When he was fourteen he began to question theological doctrines and in the course of four years abandoned successively belief in free will, immortality, and God, the last as the

result of reading John Stuart Mill’s *Autobiography*. For some time, however, Russell had metaphysical attachments that served as substitutes for religion, and it was not until the end of the first world war that he became a militant opponent of all forms of supernaturalism.

**EARLY PLATONISM AND HEGELIANISM.** Under the influence of J. M. E. McTaggart and F. H. Bradley, Russell came, in his early years at Cambridge, to believe “more or less” in the Absolute and the rest of the apparatus of British Hegelianism. “There was a curious pleasure,” Russell wrote in retrospect, “in making oneself believe that time and space are unreal, that matter is an illusion, and that the world really consists of nothing but mind.” In a “rash moment,” however, he turned “from the disciples to the Master.” G. W. F. Hegel’s remarks in the philosophy of mathematics he found “both ignorant and stupid,” and in other ways Hegel’s work appeared a “farrago of confusions.” After that Russell was converted by G. E. Moore to a “watered down” version of Plato’s theory of Ideas, regarding the subject matter of mathematics as eternal and unchanging entities whose exactness and perfection is not duplicated anywhere in the world of material objects. Eventually Russell abandoned this “mathematical mysticism” as “nonsense.” Following Ludwig Wittgenstein he came to believe “very reluctantly” that mathematics consists of tautologies. As to the timelessness of mathematics, Russell now regarded this as resulting from nothing more than that the pure mathematician is not talking about time. Aside from this, it became emotionally difficult for him to remain attached to “a world of abstraction” in the midst of the slaughter of the Great War. “All the high-flown thoughts that I had had about the abstract world of ideas,” he wrote later, “seemed to me thin and rather trivial in view of the vast suffering that surrounded me.” The nonhuman world, he added, “remained as an occasional refuge, but not as a country in which to build one’s permanent habitation.” After his abandonment of Platonism, Russell wrote, he was not able to find religious satisfaction in any philosophical doctrine that he could accept.

**PACIFISM.** Russell was interested in social questions throughout his life. He was an early member of the Fabian Society and for some time in the 1890s, under the influence of Sidney and Beatrice Webb, championed imperialism and supported the Boer War. In 1901 he had a quasi-religious experience. He became “suddenly and vividly aware of the loneliness in which most people live” and felt the need to find ways of “diminishing this tragic isolation.” In the course of a few minutes he changed his

mind about the Boer War, about harshness in the education of children and in the administration of the criminal law, as well as about fierceness in personal relations. This experience led him to write his famous essay "A Free Man's Worship" (1903). Although Russell became a pacifist right then, for another ten years or more he was preoccupied with work in mathematical logic and theory of knowledge. It was not until the war that he became passionately concerned about social issues. It is probable, he observed later, that "I should have remained mainly academic and abstract but for the War." The war, however, "shook him" out of many prejudices and made him reexamine a number of fundamental questions. He recalled:

I had watched with growing anxiety the policies of all the European Great Powers in the years before 1914, and was quite unable to accept the superficial melodramatic explanations of the catastrophe which were promulgated by all the belligerent governments. The attitude of ordinary men and women during the first months amazed me, particularly the fact that they found a kind of pleasure in the excitement. (*Selected Papers of Bertrand Russell*, p. xi)

He decided that he had been quite mistaken in believing the claims of pacifists that wars were the work of devious tyrants who forced them on reluctant populations. Although he was not then familiar with the theories of psychoanalysis, Russell concluded that the majority of human beings in our culture were filled with destructive and perverse impulses and that no scheme for reform would achieve any substantial improvement in human affairs unless the psychological structure of the average person was suitably transformed.

Russell recalls that his decision to oppose the war was made particularly difficult by his passionate love of England. Nevertheless, he had no doubt as to what he had to do. "When the war came I felt as if I heard the voice of God. I knew that it was my business to protest, however futile protest might be. My whole nature was involved. As a lover of truth, the national propaganda of all the belligerent nations sickened me. As a lover of civilisation, the return to barbarism appalled me" (*Portraits from Memory*, p. 27). Russell remarks that he never believed much tangible good would come from opposition to the war, but he felt that "for the honor of human nature," those who "were not swept off their feet" should stand their ground. He patiently argued in lectures and books that the slaughter of millions of men was not justified by any of the possible gains of a defeat of the Central Powers. Russell's pacifism was not mystical. It was not then and

had not been his contention at any time that the use of force is always wrong, that war can never possibly be justified. He maintained that *this* war in *these* circumstances was not worth all the pain and misery, and the lying of all the parties. Consistently with his general position, Russell favored the Allies during World War II on the ground that the defeat of the Nazis was essential if human life was to remain tolerable. The Kaiser's Germany by contrast was "only swashbuckling and a little absurd," allowing a good deal of freedom and democracy.

Prior to the war there had been strong pacifist sentiment in all the major Western countries, especially among the intellectuals and the powerful socialist and liberal parties. When war came only a tiny minority of these pacifists remained true to its convictions. Overwhelmed by their need to conform and in many cases by what Russell would have regarded as their own primitive impulses, many of them became the most violent jingoists. Russell was bitterly attacked for his pacifist activities not only, as one might have expected, by conservatives and professional patriots but also by many of his erstwhile friends. H. G. Wells, for example, publicly heaped abuse on Russell when he was already in trouble with the authorities. Russell's political philosophy, according to Wells, amounted to a "tepid voluntarism," and he (unlike Wells) had no right to speak for British socialism. Wells even abused Russell's work as a mathematical philosopher. Russell, he wrote, is that "awe-inspiring" man who "objected to Euclid upon grounds no one could possibly understand, in books no one could possibly read" (preface to P. H. Loyson, *The Gods in the Battle*, London, 1917).

At Cambridge, Russell's teacher and friend McTaggart led a move for his ouster. Meetings addressed by Russell were broken up by violent mobs without any police interference. Eventually he was prosecuted by the government. For writing a pamphlet on the case of a conscientious objector he was fined £100. When he would not pay the fine the government sold parts of his library, including rare books on mathematics that Russell was never able to recover. In 1918 he was sentenced to six months' imprisonment for an article in the *Tribunal*, a pacifist weekly, in which he had written that "unless peace comes soon ... the American garrison, which will by that time be occupying England and France, ... will no doubt be capable of intimidating strikers, an occupation to which the American army is accustomed when at home." In a fierce denunciation which accompanied the sentence, the magistrate, Sir John Dickinson, referred to Russell's offense as "a very despicable one" and added that Russell "seems to

have lost all sense of decency.” It should be added that as the result of the intervention of Arthur Balfour, Russell was treated with consideration while in prison—he finished there his *Introduction to Mathematical Philosophy* and began work on *The Analysis of Mind*.

**Attitude toward the Soviet Union.** Russell’s isolation was not ended with the return of peace. This was due to his failure to support the Bolshevik regime in Russia. Like many Western socialists he at first welcomed the news of the revolution, but, wanting to see things for himself, he visited Russia in 1920 and came back totally disillusioned. Some of Russell’s friends argued that any criticism of the revolution would only play into the hands of the reactionaries who wanted to reestablish the old order. After some hesitation Russell decided to publish the truth as he saw it. Russia, he later wrote, “seemed to me one vast prison in which the jailors were cruel bigots. When I found my friends applauding these men as liberators and regarding the regime that they were creating as a paradise, I wondered in a bewildered manner whether it was my friends or I that were mad.”

The little book in which he recorded his views of the Soviet Union, *The Theory and Practise of Bolshevism* (1920), was remarkable for, among other things, its prescience. Long before most Westerners had heard of Joseph Stalin, Russell predicted, point by point, the reactionary features that came to characterize the Soviet system under Stalin—its militarism and nationalism, the hostility to free art and science, its puritanism, and the gradual ascendancy of bureaucrats and sycophants over the early idealists. Russell was able to reprint the book in 1947 without a single alteration. His isolation after his return from Russia was even greater than during the war. The patriots had not yet forgiven him his opposition to the war, while the majority of his former political friends denounced him for his opposition to the Soviet regime. But Russell has never played to the galleries. As on many other occasions he acted in accordance with his favorite biblical text—“Thou shalt not follow a multitude to do evil.”

**Education and sexual morality.** Probably the most controversial of Russell’s opinions are those relating to education and sexual morality. These were closely connected with his observations of the joy people took in the fighting and killing during the war. Russell wrote that he thought he saw the inward and outward defeats that led to cruelty and admiration of violence and that these defeats were, in turn, largely the outcome of what had happened to people when they were very young. A peaceful and happy world could not be achieved without dras-

tic changes in education. In sexual matters, although not only in these, irrational prohibitions and dishonesty were exceedingly harmful. “I believe,” he wrote in *Marriage and Morals*, “that nine out of ten who have had a conventional upbringing in their early years have become in some degree incapable of a decent and sane attitude towards marriage and sex generally” (p. 249). Conventional education was judged to be at fault in a great many other ways as well. Its general tendency was to cramp creative impulses and to discourage a spirit of critical inquiry. While a certain amount of discipline is necessary, very much of the coercion traditionally employed cannot be justified. The child who is coerced “tends to respond with hatred, and if, as is usual, he is not able to give free vent to his hatred, it festers inwardly, and may sink into the unconscious with all sorts of strange consequences throughout the rest of life.”

Although puritanical moralists were or professed to be violently shocked by Russell’s views on sex and education, it is worth emphasizing that his recommendations are not extreme and that unlike his opponents he stated his position temperately and without recourse to personal abuse. Russell may be characterized as a “libertarian” in education, but he was strongly opposed to the view of other educational pioneers who played down the importance of intellectual training and encouraged originality without insisting on the acquisition of technical skill. Similarly, although he may quite fairly be called a champion of free love, it is grossly misleading to describe Russell as an advocate of “wild living.” On the contrary, he disavowed any such intentions. He wrote:

The morality which I should advocate does not consist simply in saying to grown-up people or adolescents: “follow your impulses and do as you like.” There has to be consistency in life; there has to be continuous effort directed to ends that are not immediately beneficial and not at every moment attractive; there has to be consideration for others; and there should be certain standards of rectitude. (*Marriage and Morals*, p. 243)

But this does not mean that we should be “dominated by fears which modern discoveries have made irrational.” Russell could see nothing wrong in sexual relations before marriage, and he advocated temporary, childless marriages for most university students. This, he wrote, “would afford a solution to the sexual urge neither restless nor surreptitious, neither mercenary nor casual, and of such a nature that it need not take up time which ought to be given to work” (*Education in the Modern*

*World*, pp. 119–120). It would be wrong to regard Russell as an enemy of the institution of marriage. He did indeed object to keeping a marriage going when no love is left, and, what shocked people a great deal, he remarked that a “permanent marriage” need not exclude “temporary episodes,” but he also emphatically affirmed that “marriage is the best and most important relation that can exist between two human beings ... something more serious than the pleasure of two people in each other’s company” (*Marriage and Morals*, p. 115).

Russell’s views on sexual morality featured prominently in the New York City case of 1940. When his appointment was announced, Bishop Manning of the Episcopal Church wrote an inflammatory letter to all New York City newspapers in which he denounced Russell’s subjectivism in ethics and his position on religion and morality. It was unthinkable that “a man who is a recognized propagandist against both religion and morality, and who specifically defends adultery” should be held up “before our youth as a responsible teacher of philosophy.” The bishop’s letter was the beginning of a campaign of vilification and intimidation unsurpassed in a democratic nation in recent times. The ecclesiastical journals, the Hearst press, and numerous Democratic politicians joined in the chorus of abuse. Russell was described as “the Devil’s minister to men,” as an advocate of “the nationalization of women,” as “the mastermind of free love and of hatred for parents,” and also, needless to say, as an exponent of communism.

The climax of the campaign was a taxpayer’s suit by a Mrs. Jean Kay of Brooklyn demanding that Russell’s appointment be annulled. The case was heard before Justice McGeehan, who had previously shown his notions of tolerance by trying to have a portrait of Martin Luther removed from a courthouse mural illustrating legal history. In a startling decision, which was bitterly criticized by legal experts as in many respects grossly improper, McGeehan voided Russell’s appointment on three grounds: First, Russell had not been given a competitive examination; second, he was an alien and there was no reason to suppose that the post in question could not be competently filled by an American citizen; and, finally, the appointment would establish “a chair of indecency.” Elaborate arguments were adduced in behalf of this last claim. Among other things it was maintained that Russell’s doctrines would tend to bring his students “and in some cases their parents and guardians in conflict with the Penal Law.” In some fashion not explained by the judge, Russell’s appointment would lead to “abduction” and rape. Russell’s opposition to the laws that make

homosexuality a crime was misread as advocacy of a “damnable felony ... which warrants imprisonment for not more than 20 years in New York State.” Evasive actions of the mayor of New York, Fiorello La Guardia, prevented any effective appeal against this monstrous decision, and Russell was never able to take up his position at City College. In 1950, shortly after receiving the Nobel Prize, he returned to New York to deliver the Machette Lectures at Columbia University. He received a rousing reception that those who were present were not likely to forget. It was compared with the acclaim given Voltaire in 1784 on his return to Paris, the place where he had been imprisoned and from which he had later been banished. As for McGeehan, it is safe to say that he will go down in history as a minor inquisitor who used his one brief moment in the limelight to besmirch and injure a great and honest man.

McGeehan did not pass judgment on Russell’s competence as a philosopher, but other opponents of the appointment were not so restrained. Thus, Joseph Goldstein, attorney for Mrs. Kay, described Russell as “lecherous, libidinous, lustful, venerous, erotomaniac, aphrodisiac, irreverent, narrow-minded, untruthful, and bereft of moral fiber.” After a few gratuitous lies about Russell’s private life, he concluded:

He is not a philosopher in the accepted meaning of the word; not a lover of wisdom; not a searcher after wisdom; not an explorer of that universal science which aims at the explanation of all phenomena of the universe by ultimate causes ... all his alleged doctrines which he calls philosophy are just cheap, tawdry, worn-out, patched-up fetishes and propositions, devised for the purpose of misleading the people.

In the present encyclopedia a somewhat different view is taken of the value of Russell’s philosophy. Some of his most important theories in epistemology and metaphysics will be discussed in the next section, his contributions to logic and the foundations of mathematics will be covered in the following section, and his views on ethics and religion will be dealt with in the last section. However, a number of Russell’s most interesting ideas are not at all or only briefly discussed in the present entry. Many of these are treated elsewhere in the encyclopedia.

## EPISTEMOLOGY AND METAPHYSICS

Russell exercised an influence on the course of Anglo American philosophy in the twentieth century second to that of no other individual. Yet, unlike many influential thinkers, he neither founded nor attached himself to any

definite movement. Although he wanted above all to be empirical, he always had reservations of one sort or another to the proposition that all acceptable beliefs can be derived from purely empirical premises, and although his stress on analysis as the proper philosophical method is one of the chief sources of the analytical bent that philosophy currently has in English-speaking countries, he never accepted the view that philosophy is nothing but analysis.

**EARLY REALISM.** Russell's first distinctive philosophical work was colored by a violent reaction against the absolute idealism then dominant in England, which was ultimately based on the thought of G. W. F. Hegel and whose outstanding British exponent was F. H. Bradley. According to Bradley if we try to think through the implications of any fact whatever, we will inevitably be forced to conclude that everything that there is constitutes a single, immediate unity of consciousness. In Russell's view the main weapon used to bludgeon people into submission to this result was the "doctrine of internal relations," according to which any relational fact—for example, that  $x$  is above  $y$ —is really a fact about the natures of the terms involved. This doctrine in effect refuses to take relations as ultimate.

It follows from this position that whenever  $x$  and  $y$  are related, each "enters into the nature of the other." For when  $x$  is above  $y$ , then being above  $y$  is part of the nature of  $x$  and being below  $x$  is part of the nature of  $y$ . Hence,  $y$  is part of the nature of  $x$  and  $x$  is part of the nature of  $y$ . Since everything is related to everything else in one way or another, it follows that everything else enters into the nature of any given thing, which is just another way of saying that there is no "other thing" relative to a given thing. In other words, the only thing that exists is one all-comprehensive entity. From the related principle that when we are aware of something, that something enters into the nature of the awareness or of the mind which has the awareness, it follows that it is impossible to conceive of anything which is not included within consciousness. Thus, the one all-comprehensive entity is a unity of consciousness.

Although in his youth Russell, with most of his philosophical contemporaries, was caught up in this philosophy, he and G. E. Moore became disenchanted with it shortly before the turn of the twentieth century. Russell came to hold that in sense perception we are as immediately aware of the relations between things as of the things themselves and therefore that any philosophy which denied ultimate reality to relations must be mis-

taken. Moreover, he came to think that mathematics would be impossible if we held that every relation enters into the nature of its terms; for in mathematics we must understand what our units are before we can know anything about their relations to other units. Russell therefore argued for a "doctrine of external relations," according to which relations have a reality over and above the terms they relate and do not enter into the definition of the terms they relate. This led him to a kind of philosophical atomism that thenceforth was characteristic of his philosophy. We may think of the basic core of atomism, which runs through all the shifts in Russell's later philosophizing, as being constituted by the following principles:

(1) There are nonmental facts that are what they are whether or not any mind ever becomes aware of them. This does not follow from the doctrine of external relations, but that doctrine enabled Russell to reject the idealistic argument based on the doctrine of internal relations and thus left him free to hold his native realist convictions with a good conscience.

(2) A particular proposition (for example, that my car is in the garage) can be unqualifiedly true "in isolation." This follows from the thesis that facts are "atomic" in the sense that any given fact could hold, whatever is the case with the rest of the world, together with the correspondence theory of truth—that what makes a true proposition true is its correspondence with an objective fact. Hegelians, on the other hand, had argued that since one could not adequately think about any particular fact without inflating it into the absolute totality of being, whenever one is saying something short of everything, what he is saying is not quite true in any absolute sense.

(3) An important corollary of (2) is the usefulness of analysis as a method in philosophy. If it is possible to get an adequate grasp of the parts of a totality without considering their place in the whole, then it is possible to give an illuminating account of something complex by showing how its simple parts are related to form the whole. Hegelians had argued that analysis cannot get started because we cannot understand what any part is without already seeing how it fits into the whole, which means already knowing everything about the whole. The conviction that analysis is the proper method of philosophy has remained the most prominent strand in Russell's thought.

Intoxicated by his release from idealism, Russell, as he later put it, tended to accept as objectively real anything that the absolute idealists had not succeeded in showing to be unreal. Numbers, points of space, general



properties like roundness, physical objects as they appear to sense perception, were all regarded as having an independent existence. Under the influence of Alexius Meinong this extreme realism was reinforced by an extreme form of the referential theory of meaning, the view that in order for a linguistic expression to have a meaning there must be something that it means, something to which it refers. In this stage of Russell's thought, represented most fully by *The Principles of Mathematics* and to a lesser extent by *The Problems of Philosophy*, Russell was inclined to think that the meaningfulness of the sentence "The car is in the garage" required that there be objectively existing referents not only for the words *car*, *garage*, and *in* but even for the words *the* and *is*. An objectively existing "isness" soon proved to be too much for Russell's self-proclaimed "robust sense of reality." He came to think that terms belonging to the logical framework of sentences, such as "the," "is," "or," could perform their function without each being correlated with extralinguistic referents. Nevertheless, a modified form of the referential theory of meaning continued to dominate Russell's thinking.

**LOGICAL CONSTRUCTIONISM.** Russell's decisive shift away from the full-blooded realism of *The Principles of Mathematics* came with the development of logical constructionism. The theory can be generally stated as follows. We start with a body of knowledge or supposed knowledge which we feel strongly inclined to accept but which has the following drawbacks: (1) the knowledge claims do not seem to be adequately justified, (2) there are unresolved problems about the natures of the entities involved, and (3) we feel uncomfortable about committing ourselves to the existence of such entities. If we can show that this body of knowledge could be formulated in terms of relations between simpler, more intelligible, more undeniable entities and that when so formulated there is a decisive justification for it, we will have made a philosophical advance. We will have converted the problematic to the unproblematic, the obscure to the clear, the uncertain to the certain. Russell called this technique logical constructionism because the problematic entities were said, in a possibly misleading metaphor, to be "constructed" out of the simpler ones.

**Reduction of mathematics to logic.** The technique of logical constructionism was first employed in the theory of mathematics worked out by Russell and A. N. Whitehead and published in *Principia Mathematica* (3 vols., 1910–1913). In the *Principia* the authors set out to show that all of pure mathematics can be stated in terms of logic, using no undefined terms other than those required

for logic in general—for example, implication, disjunction, class membership, and class inclusion. In the course of carrying out this reduction, various more or less problematic mathematical entities were "constructed" out of what were thought to be less problematic entities. Thus, numbers were defined as classes of classes: Zero is the class of all empty classes. The number 1 is the class of all classes each of which is such that any member is identical with any other member. The number 2 is the class of all classes each of which is such that it includes a member not identical with another member and such that any member is identical with one or the other of these. If one is puzzled about what sort of entity a number is (it does not seem to be in space or time and is not perceivable by the senses) or is uncomfortable about assuming that such queer entities exist, he will presumably be reassured by the discovery that he can think of numbers as classes of classes of familiar, unproblematic entities. Of course analogous problems may arise with respect to the entities made use of in this first reduction—for example, classes. And in fact various difficulties in doing mathematics in terms of classes led Russell to try to "construct" classes out of "prepositional functions." (See the section on logic and mathematics, below.) Starting from a given point we may well have to perform a series of reductions before we get down to maximally intelligible, indubitable entities.

**Construction of physical objects.** After *Principia Mathematica*, Russell applied the technique of logical constructionism to our knowledge of physical objects, both in physical science and in common sense. Physical theories are formulated in terms of a variety of unperceivable entities—electromagnetic fields, protons, energy quanta, forces exerted at a point, and so on. There are serious problems in the philosophy of science both about the content of our concepts of such entities and about the basis for our accepting their existence. We can try to show that such entities can be inferred from what we know about perceivable entities, but how could we get an empirical basis for a principle correlating observed and unobserved entities? Or we can try to show that unobserved entities have to be postulated in order to give an adequate explanation of observed happenings, but it seems impossible to show conclusively that no adequate explanation could be given purely in terms of observables. If we apply the constructionist principle, "Whenever possible, substitute constructions out of known entities for inferences to unknown entities," to this problem, we shall try to show that electromagnetic fields can be construed as complexes of less problematic entities related in various ways. Russell devoted a large proportion of his philosophical energy to trying to show that sci-

entific entities can be constructed out of undeniable data of perception. But it will be easier to illustrate this kind of analysis by taking ordinary physical objects like trees and buildings, for Russell thought that they raise analogous problems, although in less obvious ways.

There is a long tradition, dominant since the time of René Descartes, according to which common sense is mistaken in supposing that we directly perceive physical objects. According to this tradition what we are directly and indubitably aware of in sense perception is something private to the individual observer. There are several sources of this view, the most important of which are, first, the fact that the content of one's perception can change with, for example, changes in perspective, lighting, and physiological condition of the observer, without there being any change in the physical object which, according to common sense, one is perceiving, and, second, the fact that in dreams and hallucinations one can have experiences which are intrinsically indistinguishable from those one has when one is "really" seeing a tree, but in these cases no tree is present. In dreams and hallucinations one is really aware of something that is not a physical object and is not perceivable by anyone else. And since these experiences are intrinsically just like those in which a physical object is present, one must be perceiving these private objects in the latter cases as well. This consideration is reinforced by the first, which is designed to show that even where a physical object admittedly is involved, I am often aware of different things without the physical object's undergoing any change.

The conclusion of these arguments is that the colors, shapes, sounds, and so on, of which we are directly aware in sense perception (sense data) are private objects that must be distinguished from the entities in the physical world (if any) which we suppose ourselves to be perceiving. This conclusion inevitably gives rise to the question how, if at all, I can start from the private objects of whose existence I can be certain and show that public, physical objects like trees exist. No generally accepted solution to this problem has emerged in several centuries of discussion. Here again Russell tries to avoid the necessity for an inference by showing that the public physical objects can be construed as a complex structure of data of immediate experience. At first Russell aimed at a solipsistic reduction in which a given physical object would be constructed out of the actually experienced data of a single observer, but he soon came to lower his aspiration and to admit into the construction data experienced by others, as well as data which would have been experienced by others if they had been in a certain place. The view, then, is that a tree

can be regarded as a system of all the actual and possible sense experiences that would be regarded as figuring in perceptions of that tree. This is a form of the position known as phenomenalism, and it is subject to the difficulties to which that position is notoriously subject, particularly the apparent impossibility of specifying which experiences go into defining a particular physical object without referring to that physical object or others in the specification.

*Construction of mind.* Until about 1920 Russell was a mind-matter dualist. As we have just seen, physical objects were regarded as complex structures of data of the sort given in sense perception. Now, although the mind might be partly constituted by data which are given to "inner sense"—that is, things which are the objects of introspective awareness, such as images and feelings—it seemed to Russell, as it had to most philosophers, that in any act of awareness, be it directed to the external or to the internal world, there is in addition to the data of which one is aware a subject or self which has the experience or which performs the act of awareness. But as the spirit of logical constructionism took increasing hold of Russell, he came to feel that there was no real warrant for believing in a subject of awareness which performs acts. He became convinced that one cannot really find any such constituent of the experience; its apparent obviousness is a reflection of the grammar of the sentences in which we speak about such matters—we say "I saw a flash of light" rather than "A flash of light occurred." As it presents itself, a minimal piece of consciousness does not involve a relation between two components. It is a unitary whole. Only the flash of light is given. The "I" and the "saw" are added interpretations. If we have no real basis for accepting a subject or mind as an ultimate entity, then the logical constructionist will try to show that it can be exhibited as a complex of entities of which we are directly assured by our experience. Here Russell followed the lead of William James, who had earlier formulated a view known as neutral monism, according to which both mind and matter consisted of the data of immediate experience, the difference between them lying in the grouping of the constituents.

Thus, if I am looking at a tree the visual datum (an irregularly shaped green splotch) of which I am directly aware is both part of my mind and part of the tree. When grouped together with other experiences from this and other perspectives that would be said to be experiences of that tree, it goes to make up a tree; when grouped with other data bound together in a single conscious field, along with other data related to these by memory, it goes

to make up a mind. If this theory is acceptable, traditional puzzles about the mind-body relation are dissolved. We are faced not with two radically distinct kinds of stuff but with two different kinds of arrangement of the same elementary components. (That is, some of the components are the same. Russell considers images to be peculiar to mind.) It is in the light of this theory that one should consider Russell's notorious view that what one perceives is always his own brain. Whenever I have any sense perception whatever, I do so because a certain kind of physical activity is going on in my brain. This activity, as a physical process, is to be regarded, like all physical processes, as a construction out of the sort of data given in immediate experience. And since whatever may be the case otherwise, my brain is always active when I perceive, the data of which I am aware enter into the constitution of my brain, whatever other entities they may enter into. Hence the paradoxical view that whenever one is conscious he is aware of his own brain.

When Russell abandoned the subject of experience as an ultimate constituent of the world he rejected sense data and thenceforth spoke simply of sense experiences. But he would have represented his view more clearly by saying that he had given up belief in anything other than sense data. For in the old paradigm of *subject aware of sense data*, it was the subject exercising awareness that was abandoned. In *The Analysis of Mind* Russell set out to construct the conscious mind out of sensations and images. (Insofar as facts regarded as mental do not consist of consciousness, Russell's strategy is to give a behavioristic analysis. Thus, desire, belief, and emotions can be regarded as made up, at least in part, of dispositions to behave in one way rather than another in certain circumstances.) The results are admittedly equivocal. Russell has always been too honest to overlook glaring deficiencies in his analyses. One that has particularly bothered Russell is this: On a commonsense basis it seems clear that one must distinguish between simply having a sensation and taking that sensation as an indication of a tree, and there seems to be an important difference between simply having an image and employing that image in, for example, thinking about a forthcoming election. If this analysis of mind is to be made to work, one must give an account of the reference of perception and thought in terms of the interrelations of data. Thus, we might hold that to take a sensation as an indication of a tree is to be disposed to have the sensation of surprise if certain other sensations were to follow. But apart from difficulties about the nature of these dispositions, which are themselves neither images nor sensations, this is all extremely difficult to

work out in detail, and it is equally difficult to make sure that one has shown that it can be done.

It is clear that logical constructionism is based on a tendency opposite to that of the realism briefly espoused by Russell in his youth. Logical constructionism wields Ockham's razor with a heavy hand. We begin with those entities whose existence is indubitable because they are given in immediate experience, and we then try to show that anything we might wish to say about anything else can be stated in terms of relations between these indubitable entities. In other words, anything we want to say about something else is not really about something else. Thus, we try to represent all our knowledge as having to do with as few kinds of entities as possible, thereby reducing the possibility of error.

**LOGICAL ATOMISM.** Thus far we have concentrated on the epistemological side of logical constructionism, its concern with reducing the number of assumptions we make and with exhibiting clearly the basis for what we claim to know. But it also has a metaphysical side, although Russell wavers about this. Sometimes he talks as if his constructionism is metaphysically neutral. At such times he says that in showing that minds can be constructed out of sensations and images we do not show that there is no ultimate, irreducible subject of awareness; we show merely that everything we know about minds can be expressed without assuming the existence of such an entity. At other times, however, he claims that by showing that minds can be constructed out of sensations and images we have shown what minds really are—we have revealed their metaphysical status. And by carrying through constructions of everything that can be constructed out of simpler entities we will have developed a complete metaphysical scheme.

*Ideal language.* The most systematic presentation of this metaphysical side of logical constructionism is found in the set of lectures *The Philosophy of Logical Atomism*, which Russell gave in 1918. Here Russell makes explicit the principle on which a metaphysical interpretation of logical constructionism depends—namely, isomorphism of the structure of an ideal language and of the structure of reality. If we can determine in outline how the world would be described in an ideal language, we will have, in outline, an account of what the world is like. The restriction to an "ideal" language is essential. Since there are alternative ways of stating the same body of facts, it could not be the case that all these ways reflect the real structure of the world. In this approach to metaphysics the basic metaphysical commitment is to the identity of structure

between reality and an ideal language, and one shows one's hand metaphysically by choosing one rather than another set of criteria for an ideal language.

For Russell the most important requirement for an ideal language is an empiricist one, formulated in the "principle of acquaintance": "Every proposition which we can understand must be composed wholly of constituents with which we are acquainted." In other words, we can understand a linguistic expression only if it either refers to something we have experienced or is defined by other expressions which are so used. This principle plays a part in the constructions we have been surveying, as do the considerations we have already made explicit. That is, Russell holds not only that if physical objects were not defined in terms of sense experiences we would have no way of *knowing* anything about them but also—and even more important—we would not be able to *understand* talk about them. In logical atomism this principle is reflected in the requirement that the expressions which figure in the "atomic" sentences in terms of which everything is to be expressed must get their meaning through direct correlation with experience. They will, therefore, be names of particular sense data and terms for properties of sense data and relations between sense data. Russell is forced to exclude the logical framework of sentences from this requirement ("is," "the," etc.), but he is recurrently uneasy about this exclusion and recurrently disturbed by the question how, in that case, we can understand them.

In addition to the need for its undefined terms getting their meaning through correlation with immediately experienced items, the ideal language will have to satisfy some more strictly logical requirements. These will include the absence of vagueness and having one and only one expression for each meaning. But the most important restriction concerns the form of the basic sentences. An atomic sentence will be one that contains a single predicate or relational term and one or more than one name, the whole sentence asserting that the entity named has the indicated property ("This is white") or that the entities named stand in the indicated relation ("This is above that"). If a sentence (1) has this form, (2) contains only terms that get their meaning through correlation with experienced items, and (3) has to do with entities that cannot be analyzed into anything simpler, then it is an atomic sentence. It is clear that for Russell the sentences which satisfy these requirements will all state a minimal fact about a momentary content of sense experience.

Logical atomism can then be presented as the thesis that all knowledge can be stated in terms of atomic sen-

tences and their truth-functional compounds. A truth-functional compound of two sentences is one whose truth or falsity is a determinate function of the truth or falsity of the components. Thus, "I am leaving and you are staying" is a truth-functional compound of "I am leaving" and "You are staying." For the compound is true if and only if both its components are true. There is an empiricist motivation for maintaining this thesis. Atomic sentences, in the sense specified above, can be conclusively verified or falsified by a single experience, and as long as we are dealing only with truth-functional compounds of these no further problem can arise concerning their truth or falsity. Consider a "contrary-to-fact conditional," such as "If I had offered him more money, he would have accepted the job." As it stands this sentence is not a truth-functional compound of its constituents. For in saying it we are presupposing that both its constituents are false, yet this does not settle the question whether the whole statement is true or false. There is a corresponding puzzle about what empirical evidence would settle the question. Obviously I cannot go back in time and offer him more money and see what he will do. If we could find some way to restate this as a (very complicated) truth-functional compound of atomic sentences, it would become clear which experiences would verify or falsify it.

*Pluralism and knowledge by acquaintance.* The metaphysical correlate of this sketch of the ideal language brings together two of Russell's deepest convictions, the logical independence of particular facts (pluralism) and the dependence of knowledge on the data of immediate experience. In this view reality consists of a plurality of facts, each of which is the sort of fact which could be infallibly discerned in a single moment of experience and each of which could conceivably be what it is even if nothing else were in existence. All the familiar and seemingly relatively simple objects in the world of common sense are really extremely complicated complexes of atomic facts of these sorts.

Russell was well aware that logical atomism in this extreme form was untenable. For example, he insisted that generalizations could not be truth-functional compounds of atomic sentences. The most promising way of so construing them would be to take, for example, "All lemons are yellow" as a conjunction of a large number of atomic sentences of the form "This lemon is yellow," "That lemon is yellow," .... But as Russell points out, even if it were possible to list *all* the lemons, the conjunction would say the same thing as the original universal generalization only if we added the conjunct "and that is all the lemons there are." And this last addition is not an atomic

sentence. Moreover, Russell had doubts about so-called intensional contexts, such as “Smith believes that the White Sox will win,” where the truth or falsity of the compound is clearly independent of the truth or falsity of the components. Whether Smith has this belief does not in any way depend on whether the White Sox win. Russell has always hoped that neutral monism would help him to get out of this difficulty. If we could construct beliefs out of sensations and images we might be able to restate this fact as some truth-functional derivative of atomic sentences.

*Later doubts.* In the mid-twentieth century Russell came to have more fundamental doubts about logical atomism, including doubts concerning the very notion of a logical atom. How can we ever be sure that we are dealing with something that cannot be further analyzed into parts? How can one be sure that yellowness is an absolutely simple property? More basically, what makes a property *logically* simple? Does the fact that one can explain the word *yellow* to someone by saying “Something is yellow if it has the same color as the walls of your room” show that being the same color as the walls of your room is logically a *part* of yellowness? If so, then yellowness is not absolutely simple. If not, what does count against logical simplicity? Moreover, if there are alternative minimum vocabularies, then a simple, undefined term in one mode of formulation may turn out to be definable in another. Thus, on one systematization “pleasure” might be defined as the satisfaction of desire, whereas on a different systematization “desire” would be defined as the belief that something is pleasant. Russell gave up the belief that we can know that we have gotten down to ultimate simples and even the belief that there must be absolute simples. He became disposed to think, in more relativistic terms, of a class of things that can be taken as simple at a given stage of analysis. In those terms he still tended to fall back on sense experiences that are as apparently simple as anything we can find. Such experiences, even if not absolutely simple, can be regarded as being independent of anything except their possible components.

LATER DEVELOPMENTS. Despite Russell’s frank admissions that logical atomism does not work as a depiction of the structure of an ideally adequate language, he did not develop an alternative metaphysics. On the principle of isomorphism, if one cannot represent general statements as functions of atomic statements, then one must admit general facts as ultimate constituents of the world. This metaphysical implication did not seem to bother Russell as it once had. This is partly because he became less pre-

occupied with metaphysics in his later years and partly because the principle of isomorphism became so heavily qualified as to remove most of the cutting edge. In his major philosophical works of the 1940s, *An Inquiry Into Meaning and Truth* and *Human Knowledge*, he is more concerned with the nature of atomic facts thought of as the ultimate pieces of empirical data and the kinds of inferences required to get from these to the rest of what one wants to count as knowledge than he is with inferring a metaphysical structure from the logical form which an adequate statement of our knowledge would assume.

In these works there is a major shift in his view of the structure of atomic facts. Russell had earlier interpreted the word *this* in “This is red” as referring to a particular, something which has qualities and stands in relations but is not itself a quality or relation or set of qualities or relations. This is the traditional concept of substance as the substratum of properties, which was still alive in the realm of sense data even after physical objects and minds were no longer taken to be substances. But eventually the sense datum as substratum of properties went the way of physical objects, minds, and numbers. Here, too, Russell became convinced that there is no empirical warrant for assuming the existence of any such thing. In sense experience I am aware of a variety of qualities and their interrelations, but I am not also aware of something which *has* qualities. The bearer of qualities turns out to be the shadow of the usual grammatical form of the sentences used to report atomic facts. (There is a subject of the sentence—for example, “this”—which does not refer to any quality.) Russell’s latest position was that the subject of qualities is simply a construction out of a set of compresent qualities. Thus, in the ideal language “This is red” would be restated as “Red is compresent with ...,” where in place of the dots we have a specification of the other properties involved in that experience, for example, being round, being in the middle of the visual field, having ragged edges. It might be thought that this necessarily involves giving up the idea of absolute simples, for what takes the place of things in this view is bundles of qualities. But in this theory qualities themselves are regarded as the ultimate particulars (possibly simple) of which the world consists. Thus, in “Red is compresent with ...” “red” does not refer to a particular exemplification of redness. If we took that line we would have to suppose that there is something which distinguishes *this* exemplification from other exemplifications of just the same color, and that would have to be something as unempirical as a substratum. Instead, it is taken to refer to the color conceived as a “scattered particular,” something which can

exist in a number of different places at the same time. And such a particular might well be simple.

Russell continued to think of commonsense physical objects and the entities of physics as constructions out of entities of the sort that are given in sense experience. But he came to require less similarity to sense data in the elements of these constructions. His later view was that although all ultimate entities have basic structural similarities to sense experiences, they need not involve only qualities that are given in sense experience. They may have qualities that it is impossible for us to be aware of. This uncertainty does not carry with it any serious gap in our knowledge, since for physical science it is the structure of external events that is important. In the 1940s Russell became increasingly concerned with the principles that are required to justify inferences from sense experience to unexperienced events and complexes of unexperienced events. The simplest form this takes is, for example, the inference that my desk has continued to exist in my office throughout the night, when no one was observing it. On Russell's view this is an inference from certain sense experiences to structurally similar events spatiotemporally connected with them in certain ways. He felt that the principle of induction by simple enumeration (the more often one has observed *A* and *B* to be associated, the more it is likely that they are invariably correlated) is insufficient to justify such inferences. What is needed, he thought, is a set of assumptions having to do with spatiotemporal connections of events of like structures. In *Human Knowledge* he presents a set of such assumptions. He does not claim that they can be *known* to be true. His point is a Kantian one: We must accept these assumptions if we are to accept the inferences to unobserved events that we all do accept in the course of our daily life.

Russell's entire philosophical career was dominated by the quest for certainty. In the middle decades of the twentieth century he was driven to admit that it is less attainable than he had hoped, but nevertheless the desire to approximate it as much as possible continued to shape his thinking about knowledge and the nature of the world. Because of this desire he was continually preoccupied with the problem of how to formulate those pieces of knowledge that are rendered indubitable by experience. And because of it he consistently attempted to analyze anything that appears dubitable into constituents about which there can be no doubt. Even where he was forced to admit that inferences beyond the immediately given are inevitable, he strove to reduce the principles of such inferences to the minimum. Russell is distinguished

from other seekers after absolute certainty chiefly by the ingenuity of his constructions and by the candor with which he admits the failures of the quest.

## LOGIC AND MATHEMATICS

**REDUCTION OF MATHEMATICS TO LOGIC.** Russell's main work in logic and mathematics was concerned with the problem of bringing the two together and with the interpretation of mathematics—arithmetic in particular—as a simple extension of logic, involving no undefined ideas and no unproved propositions except purely logical ones. Russell achieved this synthesis at the beginning of the twentieth century, a little later than Gottlob Frege, but independently of him; in working it out in detail he had the collaboration of A. N. Whitehead. By current standards Russell's work lacks rigor, and in this respect it compares unfavorably with that of Frege; at an early stage, however, Russell did notice a difficulty that had escaped Frege's attention, the paradox about the self-membership of classes, which will be examined later. Because of its complexity it will be best to treat Russell's picture of the logical foundations of mathematics systematically rather than historically, with occasional comments about the actual development of his thought. We shall also separate from the outset two elements of Russell's treatment of his and other paradoxes, the theory of "types" and the theory of "orders," which Russell himself ran together, and thereby give a slightly clearer picture of his intention than his own writings immediately furnish.

*Definition of "similarity."* Russell took over from Giuseppe Peano the reduction of all other arithmetical notions to complications of the three arithmetically undefined ideas of "zero," "number," and "successor" and defined these in terms of the theory of logical relations between classes or sets. In particular, he defined a number as a class of classes with the same number of members; for example, he defined the number 2 as the class of pairs. This procedure may seem unnatural (do we really mean by "2" the class of two-membered classes?) and circular. To the charge of unnaturalness Russell's answer was that his definition (together with the definitions of addition, etc.) gives all the ordinary results ( $2 + 2 = 4$ , for example) and that for a pure mathematician this is enough; another answer can be given only after it has been made clearer what Russell means by a class. With regard to the charge of circularity, Russell defines the complex "having the same number of members," or "similarity," as he calls it, not in terms of "number" (or of his definition of "number") but in other terms altogether.

At this point some notions from the logic of relations have to be introduced. A relation is said to be one to one if whatever has that relation to anything has it to one thing only and if whatever has anything standing in that relation to it has one thing only standing in that relation to it. (In strictly monogamous countries, “husband of” is a one-to-one relation in this sense.) Here the phrase “one only” does not presuppose the notion of the number 1. The sentence “ $x$  stands in the relation  $R$  to one thing only” means “For some  $y$ , whatever  $x$  stands in the relation  $R$  to is identical with  $y$ .” The domain of a relation is the set of objects that stand in that relation to anything (the domain of “husband of” is the class of all husbands); the relation’s converse domain is the set of all objects to which anything stands in that relation (the converse domain of “husband of” is the class of individuals that have husbands—that is, the class of wives). A class  $A$  is similar to (that is, has the same number of members as) another class if there is some one-to-one relation of which the first class is the domain and the second the converse domain.

One can see that in a monogamous country the class of husbands will be similar in this sense to the class of wives, but one might think that two sets of objects could have the same number of members without there being any relation at all that pairs them off in the way that “husband of” does in our example. This, however, is a mistake when the term *relation* is understood as widely as it is by Russell. A relation in Russell’s sense is, roughly, anything that can be expressed by a sentence with two gaps in it where names might go, and this covers not only obvious relating expressions like “\_\_\_\_\_ shaves ( )” or “\_\_\_\_\_ is the husband of ( )” but also ones like “Either \_\_\_\_\_ is identical with  $A$ , or  $B$  is identical with ( ).” Take any set of two objects  $C$  and  $D$ . The relation “Either \_\_\_\_\_ is identical with  $A$  and ( ) with  $C$ , or \_\_\_\_\_ is identical with  $B$  and ( ) with  $D$ ” (where all dashes must be replaced by the same name, and similarly with the bracketed blanks) will be a one-to-one relation in which  $A$  stands to  $C$  alone and  $B$  to  $D$  alone and in which  $C$  has  $A$  alone standing to it and  $D$  has  $B$  alone—that is, it will be a one-to-one relation of which the class with  $A$  and  $B$  as sole members is the domain and the class with  $C$  and  $D$  as sole members the converse domain; there are analogous relations in the case of larger classes. (Where the classes are infinitely large these relations will not be expressible in a language with only finite expressions, and perhaps that means that they will not be expressible in any language. Some philosophers would regard this as a serious difficulty; others would not.)

*Axiom of infinity.* Similarity, then, or having-the-same-number-of-members, is defined in terms of notions from the logic of relations: one to one, domain, and converse domain. The number-of-members of a given class is the class of classes similar to it, and a class of classes is a number (strictly, a cardinal number) if there is some class of which it is the number-of-members. This last step gives rise to another difficulty: Suppose there are (as there might well be) no more than a certain number  $n$  of objects in the universe. Then there will be no classes with more than  $n$  members and so, by the above definition, no cardinal numbers greater than  $n$ . This makes a great part of arithmetic (for example, the principle that every number has a successor different from itself) subject to the hypothesis (sometimes called the axiom of infinity) that there are an infinite number of objects.

Russell came to accept this last consequence of his definitions, but at an earlier stage he had thought that the axiom of infinity was provable, as follows: If we assume that every property demarcates a class, we must admit that some classes are empty (have no members), for example, there are no objects not identical with themselves. (The number 0 is precisely the class of classes with no members.) Thus, even if the universe contains no ordinary objects at all, there will still be at least one object of a more abstract sort, the universe itself considered as an empty class. And if there is this object there will also be two further objects of a still more abstract sort: the class of classes that has the first empty class as its one member and the empty class of classes. That makes three objects, call them  $A$ ,  $B$ , and  $C$ . In addition to these there will be four classes of classes of classes—the class with  $B$  as its sole member, that with  $C$  as its sole member, that containing both  $B$  and  $C$  as members, and the empty class of classes of classes. And so on ad infinitum.

*Russell paradox and the theory of types.* Russell was led to abandon the above demonstration (which, as he said, has anyway “an air of hocus-pocus about it”) by his discovery of the paradox of self-membership, mentioned earlier. If we can concoct classes with some members that are themselves classes, some that are classes of classes, and so on as we please (if, in other words, we can treat classes, classes of classes, etc., as so many sorts of classifiable “objects”), we can, it seems, argue as follows: The most obvious classes do not contain themselves as members—for example, the class of men is not itself a man and so is not itself a member of the class of men (that is, of itself). On the other hand, the class of non-men *is* a non-man (is one of the things that are not men) and thus *is* a member of itself. We can therefore divide classes into two broad

classes of classes—the class of classes that are members of themselves and the class of classes that are not. Now take the class of classes that are not members of themselves: Is it a member of itself or not? If it is, it must possess the defining property of this class to which ex hypothesi it belongs—that is, it must be not-a-member-of-itself. (Thus, if it is a member of itself, it is not a member of itself.) And if it is not a member of itself, ipso facto it possesses its own defining property and so *is* a member of itself. (If it is not, it is.) Let  $p$  be the proposition that our class is a member of itself; it follows even from the attempt to deny it, so it *must* be true—but it entails its own denial, so it *must* be false. There is clearly something wrong here.

Russell thought the error lay in treating a class seriously as an object. Perhaps it is an object in a sense, but not in the same sense in which genuine individuals are objects—and classes of classes are different again. They are, as he put it, of different “logical type.” In particular, in an intelligible sentence you cannot replace an individual name by a class name or a class name by the name of a class of classes, or vice versa, and still have the sentence make sense. If “Russell is dead” makes sense, “The class of men is dead” does not, and if “The class of men is three-membered” makes sense (even if false), “Russell is three-membered” does not. And where a sentence makes no sense (as opposed to being merely false), its denial makes no sense either. Since “The individual  $I$  is a member of the class-of-individuals  $C$ ” makes sense, “The class-of-individuals  $C$  is a member of the class-of-individuals  $C$ ” does not and neither does “The class-of-individuals  $C$  is not a member of the class-of-individuals  $C$ ”—and so on at higher points in the hierarchy. This being granted, the paradox with which we began simply cannot be intelligibly formulated and thus disappears from the system.

At this point it would be wise to remove a possible source of confusion. The relation of class membership is different from the relation of class inclusion. One class is included in another if all the members of the former are members of the latter; for example, the class of men is included in the class of animals—all men are animals. But the class of men is not a member of the class of animals; that is, the class of men is not an animal (or, more strictly, “The class of men is an animal” is nonsense). The class of men is a member, rather, of the class of classes-of-animals—it is a class of animals. And the class of classes of animals is included in (but is not a member of) the class of classes of living things—any class of animals, in other words, is a class of living things. Inclusion thus relates classes of the same logical type; membership, on the other

hand, relates an entity with another entity of the logical type one above its own. The membership of an individual in a class of individuals is membership in a sense different from the membership of a class of individuals in a class of classes, and similarly for inclusion—there is a hierarchy not only of classes but also of membership and inclusion relations.

All this, besides solving a technical problem, is not without some attraction for philosophical common sense. Even apart from paradoxes it seems an artificial “multiplication of entities” to suppose that in addition to the individual objects which form the members of the lowest type of classes there are classes, classes of classes, and so on, and Russell devoted some attention to the problem of showing how what appears to be talk about these rather strange objects is in reality just more and more oblique talk about quite ordinary ones. To see just how he shows this it is necessary to look more closely at what might be called his “straight” language, into which this talk of classes, etc., does not enter and into which, once this talk *has* been introduced, it can always be “translated back.”

LOGIC. From what has been said so far, it is clear that the “logic” to which Russell reduced arithmetic covered, implicitly or explicitly, such subjects as class membership and class inclusion, identity, and some sort of theory of relations. This is that part of logic that we first encounter when we work back to logic from arithmetic. We must now try and work forward to the same point from the fundamentals of logic.

Russell thought of logic as being at bottom “the theory of implication” (to quote the title of one of his early papers). And from the first he considered it important to distinguish implication from inference. He objected to the view that logic is primarily about thinking—conception, judgment, and inference, as some of the traditional logic texts put it. The connection of logic with inference is rather that logic is concerned with that in the real world that makes inference justified, and this is implication. “Where we validly infer one proposition from another,” he wrote in 1903, “we do so in virtue of a relation which holds between the two propositions whether we perceive it or not: the mind, in fact, is as purely receptive in inference as common sense supposes it to be in perception of sensible objects” (*Principles of Mathematics*, p. 33).

**Material implication.** Even in Russell’s purely objective, nonpsychological sense “implication” is ambiguous. Implication may be a relation between complete propositions, in which case it is called “material” implication and



holds whenever it is not the case that the implying proposition is true and the implied proposition false. Before enlarging and commenting upon this account, certain grammatical and metaphysical clarifications are in order. Russell originally believed that sentences symbolized abstract objects called “propositions” and that material implication was a relation between these objects in exactly the same sense that marriage might be a relation between two people. He later dropped this view and regarded propositions, like classes, as mere “logical constructions,” but he still used the old forms of words (as being, no doubt, accurate enough for practical purposes). In particular, the partly symbolic form “ $p$  implies  $q$ ” (or “ $p$  materially implies  $q$ ”) freely occurs in all his writings, and we ought to be clear about what he means by it. Generally it is simply a variant of “If  $p$  then  $q$ ,” or completely symbolically “ $p \supset q$ ” (“ $p$  hook  $q$ ”), where the phrase “If \_\_\_\_\_ then ( )”—or the hook—is not a transitive verb expressing a relation between objects but a conjunction, or, as we now say, a “sentential connective.” “If  $p$  then  $q$ ” is thus not a statement about two objects symbolized by “ $p$ ” and “ $q$ ” but rather a complex statement about whatever the statements represented by “ $p$ ” and “ $q$ ” are about. For example, “If James is going to come, John will stay away” is not about two objects symbolized by “James is going to come” and “John will stay away,” nor is it about these subordinate sentences themselves; rather, it links these two sentences to make a more complex statement about James and John. And if we say “That James is going to come implies that John will stay away,” this is just a verbal variant of “If James is going to come then John will stay away”; that is, the linking expression “That \_\_\_\_\_ implies that ( )” has the same meaning as the conjunction “If \_\_\_\_\_ then ( ).” The general form “That  $p$  implies that  $q$ ” thus has the same sense as the form “If  $p$  then  $q$ ” or “ $p \supset q$ ,” and Russell’s “ $p$  implies  $q$ ” is thus just a loose way of saying “That  $p$  implies that  $q$ .” In a similar way Russell often uses “ $p$  is true” and “ $p$  is false” as variants of “It is the case (is true) that  $p$ ” and “It is not the case (is false) that  $p$ ”; although sometimes he may really be talking about sentences in such a way that the sentence “John will stay away” may be described as true if and only if John will stay away and as false if and only if he will not, and the sentence “James is going to come” may be said to “imply” the sentence “John will stay away” if and only if the sentence “If James is going to come then John will stay away” is true.

The assertion that an implication is true if and only if it is not the case that the implying statement (antecedent) is true and the implied statement (consequent) false is not intended as a definition of the form “If

$p$  then  $q$ .” It is simply an informal attempt to fix our attention on the relation (or quasi relation) that Russell intends. In his earliest works, like Frege and C. S. Peirce before him, Russell took this relation to be undefinable, and “the discussion of undefinables—which forms the chief part of philosophical logic—is the endeavour to see clearly, and to make others see clearly, the entities concerned, in order that the mind may have that kind of acquaintance with them which it has with redness or the taste of a pineapple” (*Principles of Mathematics*, 1st ed., preface; 2nd ed., p. xv). Later he preferred to take as undefined the conjunction “or” and the negative prefix “it is not the case that” (or just “not”) and to define “If  $p$  then  $q$ ” as an abbreviation of “Either not  $p$  or  $q$ ”; later still he followed H. M. Sheffer and Jean Nicod in using the stroke form “ $p | q$ ” (which is true if and only if the component statements are not both true) and defined “if,” “not,” and “or” in terms of it. But for Russell the central part of logic has always been the study of implication, whether taken as undefined or not.

Since the form “If  $p$  then  $q$ ” as understood by Russell is true as long as it is not the case that the antecedent is true and the consequent false, it is automatically true if the antecedent is false (for then it is not the case that the antecedent is true and thus not the case that the-antecedent-is-true-and-the-consequent-false) or the consequent true (for then it is not the case that the consequent is false and thus not the case that the-antecedent-is-true-and-the-consequent-false). In other words, a false proposition materially implies, and a true one is materially implied by, any proposition whatever. But implication is supposed by Russell to justify inference, and the mere fact that “Grass is pink” is false would not seem to justify us in inferring the 25th proposition of Euclid from it, and the mere fact that Euclid’s proposition is true would not seem to justify us in inferring it from “Grass is green”—geometry would be much easier if we could do this. Russell’s explanation is that the first of these inferences cannot be performed because we cannot get it started (the premise not being true) and that the second inference is justified but we cannot know it to be so unless we already know the conclusion, so that we will not need it. In other words, “Infer a true proposition from anything at all” is a rule with no practical use, but this does not make it logically wrong.

#### *Formal implication and propositional functions.*

Implications are of practical use when we know their truth without knowing either the falsehood of their antecedents or the truth of their consequents, and this happens most often when a material implication is an

instance or particularization of an implication in the second of Russell's senses, a "formal" implication.

Formal implication is not (to use Russell's "realistic" language) a relation between propositions but one between what he calls "propositional functions." One might say roughly that formal implication is a relation between properties and that one property formally implies another if it is never present without the other; for example, being human formally implies being mortal (nothing is human without being mortal). Formal implication is clearly involved in the notion of class inclusion—*A* is included in *B* if being a member of *A* formally implies being a member of *B*. But the notion of a propositional function is wider than that of a property. It is what is meant by an "open sentence," a sentence in which some expression—say, a name—has been replaced by a variable. "Socrates is a man" expresses a proposition; "x is a man" expresses a propositional function. Sometimes, more simply, Russell uses the term *propositional function* for the open sentence itself. And the proposition that Socrates is a man may be said to be the *value* of the propositional function "x is a man" for the value "Socrates" of the argument *x*. The propositional function "x is a man" formally implies "x is mortal" if *x*'s being a man materially implies that *x* is mortal whatever *x* may be—that is, if we have "For any *x*, if *x* is a man then *x* is mortal." Russell writes this sort of implication as " $\phi x \supset \psi x$ ." At one stage he treated this notion, for systematic purposes, as undefined, but even then he regarded it as complex in meaning, being built up from material implication together with the prefix "for any *x*," called a quantifier. Writing this last as "*(x)*," we may spell out the sense of a formal implication by writing it as " $(x) : \phi x \supset \psi x$ ." It should be noted that whereas a propositional function is not a proposition, a formal implication between such functions is a proposition. The propositional function "x is a man" is neither true nor false; only its various values are true or false. But "For any *x*, if *x* is human then *x* is mortal" is as it were complete and is as it happens true. The quantifier is said here to "bind" the variable *x*, or, in the terminology Russell took over from Peano, *x* is in this context not a "real" but an "apparent" variable.

A propositional function may also have more than one expression in a proposition replaced by a variable, as in "x shaves y," "x gives y to z," and "If x shaves y then x does not shave z." In such cases the function corresponds to a relation (two-termed or many-termed) rather than to a property, and such functions may again be linked by formal implication, as in "For any *x* and *y*, if *x* is a child of *y* then *x* detests *y*"—that is, "All children detest their par-

ents." Symbolically, we have here the form " $\phi xy \supset_{x,y} \psi xy$ ," or " $(x,y) : \phi xy \supset \psi xy$ ." Again, formal implication may link a propositional function and a complete proposition, as in "If anything is in that box I'm very much mistaken," which is of the form "For any *x*, if  $\phi x$  then *p*" or " $\phi x \supset_x p$ ." Moreover, the expression whose place is taken in a propositional function by a variable need not be a name. It might, for example, be a sentence—"If *p* then *q*" is a propositional function of which "If James is going to be there then John will not come" is the value when "James is going to be there" is the value of the argument *p* and "John will not come" the value of the argument *q*. If we prefix quantifiers to forms of this sort we obtain further formal implications, including the laws of propositional logic themselves—for example, "For any *p*, *q*, and *r*, if *p* implies *q* then if *q* implies *r*, *p* implies *r*," which may be written " $(p,q,r) : (p \supset q) \supset ((q \supset r) \supset (p \supset r))$ " or " $(p \supset q) \supset_{p,q,r} ((q \supset r) \supset (p \supset r))$ ."

A further case of special interest is that in which a variable replaces a verb or equivalent expression, as in " $\phi(\text{Socrates})$ ," where  $\phi$  stands indifferently for "is a man," "smokes," "is running," etc. With appropriate quantifiers this function will yield such formal implications as " $(\phi) : \phi a \supset \phi b$ ," " $\phi a \supset_{\phi} \phi b$ " (roughly, "Whatever *a* does, *b* does," or "Whatever goes for *a* goes for *b*"). However, Russell says not that " $\phi(\text{Socrates})$ " and "If  $\phi(\text{Socrates})$  then  $\phi(\text{Plato})$ " are functions of the verb or predicate  $\phi$  but that they are functions of the *function*  $\phi x$  or, as he writes it in this type of context, the function  $\phi \hat{x}$  (the significance of this accenting or "capping" will be indicated later). His aim here is in part to bring out what Peirce and Frege called the "unsaturatedness" of verbs: The function of verbs can be understood only in relation to names and sentences; we use verbs to *make statements* about objects, not to name a special sort of object. The additional associated variables also enable one to represent unambiguously such complexes as "shaving oneself"—if  $\phi$  is "shaves," shaving oneself is  $\phi \hat{x} \hat{x}$ , as opposed to simply shaving ( $\phi \hat{x} \hat{y}$ ). But Russell was hampered by not having the word *functor* to designate what makes a function out of its argument; it is more natural to speak of "Socrates is a man" as a propositional function of "Socrates," of "x is a man" as the same propositional function of "x," and of "is a man" as the functor which forms this function in both cases than to speak of " $\hat{x}$  is a man" as a propositional function and to treat it in practice as a functor (Frege and W. E. Johnson were more accurate here, although they, too, lacked the term *functor*).

This part of Russell's "philosophical grammar" can now be set out fairly straightforwardly: Sentences may be

built out of other units in various ways—out of other sentences by connectives, as in “ $p \supset q$ ” and out of names by verbs, as in “ $\phi x$ ,” “ $\phi xy$ ,” and “ $\phi x \supset \psi x$ ” (which may be conceived of as constructed out of the subsentences “ $\phi x$ ” and “ $\psi x$ ” by the connective “ $\supset$ ” or out of the name “ $x$ ” by the complex verb “ $\phi x \supset \psi x$ ,” “ $\psi$ ’s-if-it- $\phi$ ’s”). The rest of the hierarchy goes on from here—there are, for example, functors that form sentences out of verbs (that is, out of functors that form sentences out of names) and functors that form sentences out of these again, and so on ad infinitum. Functors may require one or more than one argument to make a sentence (the difference between “is a man” and “shaves,” in the transitive sense), and when more arguments than one are required they may or may not be of the same type (for example, “If  $x$  is a man then  $p$ ” requires a name and sentence).

**Quantification.** Of functors that form sentences from verbs, the most important are quantifications, such as “ $(x)\phi x$ ” (which makes a sentence out of the verb whose place in the sentence is kept by “ $\phi$ ”), represented in English by such words as “everything,” “Everything” is, or is constructed out of, the universal quantifier; there are many other quantifiers. Russell distinguished one other basic quantifier, “something.” “Something is a man” expands in his language to “For some  $x$ ,  $x$  is a man,” or symbolically  $(\exists x)(x \text{ is a man})$ .

Given the quantifier “something” and negation we can construct the complex “It is not the case that (for some  $x$  ( $x$  is a man))” or “For no  $x$  is  $x$  a man.” Here we have the philosophical beginnings of the number series. The number 0 makes its appearance as part of a quantification, for we could write the preceding form as  $(0x)(x \text{ is a man})$ . And the series can be continued. “Some” means “At least one, and “At most one thing is a man” is “For some,  $x$ , if anything is human it is identical with  $x$ ”—that is, “For some  $x$ : for any  $y$ , if  $y$  is human  $y$  is identical with  $x$ .” The combination of “At least one thing is a man” with “At most one thing is a man” gives us “Exactly one thing is a man” that is, “ $(1x)(x \text{ is a man})$ .” “ $(2x)(x \text{ is a man})$ ” is, similarly, “At least two things are men, and at most two things are men”—that is, “(For some  $x$  and for some  $y$ ,  $x$  is a man,  $y$  is a man, and  $y$  is not identical with  $x$ ) and (for some  $x$  and for some  $y$ : for any  $z$ , if  $z$  is a man  $z$  is either identical with  $x$  or identical with  $y$ ),” Apparent occurrences of numbers as objects can be analyzed away in terms of this primary sense; “1 and 1 is 2” for instance, becomes “For any  $\phi$  and for any  $\psi$ , if exactly one thing  $\phi$ ’s, exactly one thing  $\psi$ ’s, and nothing does both, then exactly two things either- $\phi$ -or- $\psi$ .” Numbers are inseparable components of functors of functors of names, or, as Russell

would say, functions of functions, but the naturalness of this analysis is disguised in his own work by the fact that before he brings arithmetic into the picture he introduces the language of classes and defines numbers in terms of classes. (The notation “ $(0x)\phi x$ ,” etc., is not Russell’s.)

**DESCRIPTIONS.** Before going on to Russell’s discussion of classes, we should note that “ $(x)\phi x$ ,” “ $(\exists x)\phi x$ ,” and also “ $(0x)\phi x$ ,” “ $(2x)\phi x$ ,” and so on, are *functions* of functions of names, not *arguments* of such functions—that is, they are not names. “Something,” “nothing,” “exactly one thing,” etc., are not names, although, like names, they go with verbs to make sentences. They go, so to speak, on the other side of verbs: They “govern” the verbs; the verbs do not govern them. And although Russell’s hierarchy of types of functors or “functions” provides innumerable ways of constructing sentences (and so of constructing functions), it provides no way of constructing genuine names. It is of the essence of the expressions represented by Russell’s variables of lowest type ( $x$ ,  $y$ ,  $z$ , etc.)—that is, individual names—that they are logically structureless; they pick out individuals, and that is all. But in common speech and in mathematics we do seem to construct names, or at least ways of designating objects, out of expressions of other types: For example, “the man who broke the bank at Monte Carlo” seems to function as a name, yet it seems to be constructed from the verb “broke the bank at Monte Carlo.” On Russell’s view this appearance is illusory, and sentences in which such apparent names occur can always be replaced by paraphrases expressed entirely in Russell’s language of structureless names, functions of functions, etc. However, he regarded it as useful for logical symbolism to reproduce at this point, although with greater precision, some of the devices of common speech and to have, as it were, a secondary language imposed on the primary one.

(Some account of Russell’s handling of descriptions—that is, expressions of the form “The so-and-so”—and of other points raised below is given in the entry Existence.) “There  $\phi$ -er,” or “The thing that  $\phi$ ’s,” when it occurs as the apparent subject of a further verb—that is, in a context of the form “The  $\phi$ -er  $\psi$ ’s”—is in reality a functor, in some way like a quantifying expression, of which the verbs “ $\phi$ ’s” and “ $\psi$ ’s” are arguments; in fact “The thing that  $\phi$ ’s  $\psi$ ’s” amounts precisely to “Exactly one thing  $\phi$ ’s, and whatever  $\phi$ ’s  $\psi$ ’s,” whose first component has been analyzed above and whose second component is a simple formal implication. Expressions of this kind are especially important in mathematics when the contained functor  $\phi$  is relational in form, as in “The  $\phi$ -er of  $y$ ”—that is, “The thing that  $\phi$ ’s  $y$ ,” “The square root of  $y$ ” (the num-

ber that yields  $y$  when multiplied by itself) is such an expression. Russell called expressions of this kind “descriptive Functions.” They include most “functions” in the ordinary mathematical sense. It is a little inaccurate, of course, to use name symbols like “ $y$ ” for numbers, which on Russell’s view are not genuine individuals, but once the devices that yield class language and number language have been worked out, Russell’s analysis of descriptive functions can be reproduced at the new level in a transposed form. This language of classes and numbers, to which we shall now turn, is itself a case of a secondary language containing apparent names (like “The class of persons that shave themselves”) that disappear from the primary-language paraphrase.

**CLASSES, FUNCTIONS, AND PROPERTIES.** Russell represented the form “The class of things that  $\varphi$ ” as “ $\hat{x}(\varphi x)$ ”—usually read as “the  $x$ ’s such that  $\varphi x$ —and represented “ $y$  is a member of the class of  $\varphi$ -ers” as “ $y \in \hat{x}(\varphi x)$ .” Alternatively we may read “ $\hat{x}(\varphi x)$ ” simply as “ $\varphi$ -er” and “ $y \in \hat{x}(\varphi x)$ ” as “ $y$  is a  $\varphi$ -er.” The expression “ $y$  is a  $\varphi$ -er” is true if and only if  $y \varphi$ ’s. One can in fact simply define “ $y \in \hat{x}(\varphi x)$ ” as “ $\varphi y$ .” Given this definition, other concepts associated with class theory are easily introduced. For example, as noted earlier, “The class of  $\varphi$ -ers is included in the class of  $\psi$ -ers” amounts to the formal implication “For any  $x$ , if  $x$  is a  $\varphi$ -er then  $x$  is a  $\psi$ -er.”

Classes of classes are related to functions of functions as classes are related to functions. To say that a given class— $\hat{x}(\varphi x)$ , for example—is a member of the class of two-membered classes (or, as Russell would write it, “ $\hat{x}(\varphi x) \in 2$ ”) is just to say that exactly two things  $\varphi$ —i.e., the class of classes that Russell identifies with the number 2 is just the correlate in the class hierarchy of the function of functions  $(2x)\varphi x$ .

**Counting classes.** There are two difficulties in Russell’s views concerning classes. One is that classes, and, for that matter, numbers, can themselves be counted, as can individuals, but a number of classes would have to be not a class of classes but a class of classes of classes, and a number of numbers would similarly have to be a class of classes of classes of classes. This means that when we say “The number of numbers between 2 and 5 is 2,” the first “2” has a sense quite different (belongs to a place quite different in the type hierarchy) from the second; and this seems a little implausible. Russell at this point is content to speak of the “systematic ambiguity” of the key expressions of his symbolic language. Given the proof of “ $1 + 1 = 2$ ,” for instance, considered as a statement about numbers of individuals, an analogous proof can always be

constructed for the analogous statements about numbers of classes, numbers of numbers, etc., so that in practice it does not matter at which place in the type hierarchy we are working, provided we keep the types going up in order.

Ludwik Borkowski has suggested what may be a better solution: Suppose we always express quantification by a sign followed by a variable; for Russell’s “ $(x)$ ” we might put “ $(\forall x)$ ,” by analogy with “ $(\exists x)$ .” We might then use the term *quantifier* not for this expression as a whole but for the initial sign, which can then be described as a functor that constructs a sentence out of a variable followed by a sentence, usually an “open” sentence in which the variable just mentioned occurs. We might then say that the initial sign “ $\forall$ ” or “ $\exists$ ”—or in the case of numerical quantifiers “0” or “1” or “2,” etc.—is of the same logical type whatever the type of the variable that comes between it and the sentence following it. For counting properties (and, therefore, classes), we would have prefixes like “ $(2\varphi)$ ”—for example, “ $(2\varphi)\varphi(\text{Socrates})$ ” would mean “Socrates has exactly two properties” or, better, “Exactly two things are true of Socrates”; and “ $(2\varphi)$ ” is different from “ $(2x)$ ,” but the “2” is exactly the same in both contexts.

**Counting functions.** The other difficulty in Russell’s theory is that classes dissolve into functions, but we do not count classes and functions in quite the same way. We would say, for example, that any two-membered class has four subclasses, in the sense that there are four ways of selecting members from such a class (both members, the first only, the second only, and neither). The corresponding theorem about functions would seem to be this: If exactly two things  $\varphi$ , then for exactly four  $\psi$ ’s, whatever  $\psi$ ’s  $\varphi$ ’s. But in fact there will always be vastly more than four  $\psi$  meeting this condition. Suppose, for example, that there are just two men in a room—i.e.,  $(2x)(x \text{ is a man in the room})$ —and that one of them wears spectacles, spats, spotted socks, a red tie, and striped trousers; this much alone gives us five  $\psi$ ’s (namely, “\_\_\_\_\_ is a man in the room wearing spectacles,” “\_\_\_\_\_ is a man in the room wearing spats,” etc.), such that whatever  $\psi$ ’s is a man in the room. The key point here is simply that we count classes as being the same when they have the same members, but we do not count propositional functions as being the same merely because they are satisfied by the same arguments, and all the numerical concepts that are built up from the concept of identity must be similarly adjusted. For instance: “At most one class is a sub-class of  $\hat{x}(\varphi x)$ ” does not mean “For some  $\psi$ : for any  $\chi$ , if  $\hat{x}(\chi x)$  is a subclass of  $\hat{x}(\varphi x)$ , then  $\chi$ -ing is the same as  $\psi$ -ing,” but rather it means “For some  $\psi$ : for any  $\chi$ , if  $\hat{x}(\chi x)$  is a sub-

class of  $\hat{x}(\varphi x)$ , then whatever  $\chi$ 's  $\psi$ 's and whatever  $\psi$ 's  $\chi$ 's." It is the same when we move up a type and count numbers themselves. If we write " $(0x)(\varphi x)$ " for "It is not the case that (for some  $x$ ,  $\varphi x$ )" and " $(0'x)(\varphi x)$ " for "For any  $x$ , if  $x \varphi s$  then  $x$  is not identical with itself," we may say that these are different functions of functions—but whatever function either of them applies to the other applies to also; thus, they determine a single class of classes and a single "number," 0. The class and number language that Russell superimposes on his basic one is such that this is the way these quasi entities are counted.

**Extensionality.** One very radical way of simplifying this whole problem (one that Russell has considered from time to time) is to say that functions (properties, relations, etc.) are to be counted in just the same way that classes are; that is, that if  $\varphi \hat{x}$  and  $\psi \hat{x}$  characterize precisely the same objects (are formally equivalent), they are the same function. This is called the principle or law of extensionality; it in effect simply identifies a function with its "extension"—that is, with the class that it determines. The objection to this principle is simply its extreme implausibility in particular cases. For example, it seems obvious that even when two individuals and these two only are the men in a certain room wearing spats and the men in that room wearing spectacles, being a man in the room with spats is something different from being a man in it with spectacles.

**Quine's criticism.** Logicians such as W. V. Quine, following Ernst Zermelo and John von Neumann, have developed systems in which classes, classes of classes, and so on, are treated not as logical constructions but as genuine objects, and Russell's paradox is dealt with not by saying that " $x$  is (is not) a member of  $x$ " is meaningless but by denying that " $x\varphi$ 's" always implies that  $x$  is a member of the class of  $\varphi$ -ers. This account runs into difficulty when we try to handle certain nonmathematical properties of these supposed objects. Russell's view seems to have the advantage of not unnecessarily "multiplying entities," but Quine argues that Russell succeeds in dispensing with classes only by making genuine objects of properties or functions. This is said on the ground that in the course of his treatment of classes and numbers Russell is compelled to quantify over predicate variables—that is, to employ such quantifiers as " $(\exists \varphi)$ " (for example, in defining "Exactly as many things  $\psi$  as  $\chi$  as "For some relation,  $\varphi$  whatever  $\varphi$ 's anything  $\psi$ 's and vice versa, whatever is  $\varphi$ 'd by anything  $\chi$ 's and vice versa, and whatever  $\varphi$ 's or is  $\varphi$ 'd by anything  $\varphi$ 's or is  $\varphi$ 'd that thing only"). This, Quine says, is to make properties and relations (like

$\varphi$ -ing) the "values of bound variables," and to do this is to treat them as existing.

This amounts to saying that to generalize an expression by quantifying over it is ipso facto to make it a name of an object; but this claim may be contested. We do not elucidate "He must have killed him somehow" by translating it "There must be some way in which he killed him" (which, taken literally, suggests that there are objects called "ways") but rather vice versa: We understand "somehow" directly as a generalization of qualifications like "with a knife," and the "way" line of talk is merely a variant of this. Even "something" is often to be understood as a generalized adjective rather than as a generalized individual name—for example, when I say "I am something that Jones is not—logical." It seems more plausible to interpret "I have something that Jones has not—logicality" as a verbal variant of the preceding sentence than to say that the latter alone brings out what I am really doing. And the logical rules for such higher-order quantifications are simple—we proceed from the specific case to the generalization, from "I am logical and Jones is not" to "For some  $A$ , I am  $A$  and Jones is not  $A$ ," exactly as we do from "I am logical but not intelligent" to "For some individual  $x$ ,  $x$  is logical but not intelligent."

**Elimination of abstract terms.** Russell might more plausibly be said to "hypostatize" or "reify" abstractions on the ground that there are some contexts from which it seems impossible to eliminate from his basic language his symbols for "abstracts," that is,  $\varphi x$ , etc. This part of his system is developed more tidily in Alonzo Church's calculus of  $\lambda$ -conversion, in which the property of  $\varphi$ -ing is represented not by " $\varphi \hat{x}$ " but by " $\lambda x \varphi x$ ," and of " $\psi$ -ing if one  $\varphi$ 's" by " $\lambda x . \varphi x \supset \psi x$ ." The basic rule of this calculus is that the application of  $\lambda x \varphi x$  to an object  $a$ , symbolized by  $(\lambda x \varphi x)a$ , is equivalent to the plain  $\varphi a$ , and similarly  $(\lambda x . \varphi x \supset \psi x)a$  is equivalent to  $\varphi a \supset \psi a$ . And where we have a function of functions  $f$ , we can in general similarly replace  $f(\lambda x \varphi x)$  by  $f(\varphi)$ —but not always. For instance, it is an obvious law that any such function  $f$  which holds for any  $\varphi$  whatever will hold for  $\chi$ -ing-if-one- $\psi$ 's, as in formula  $F$ :

$$(f) : (\varphi) f(\varphi) . \supset . f(\lambda x . \psi x \supset \chi x).$$

Here the expression with  $\lambda$  seems uneliminable. We cannot replace it with " $\psi \supset \chi$ ," for this is meaningless—the hook joins sentence forms, not predicate forms. Where we have a specific  $f$  the elimination is again possible; for example, if  $f$  is the function "applying to exactly two objects," then  $f(\lambda x . \psi x \supset \chi x)$  will amount to  $(2y) : (\lambda x . \psi x \supset \chi x)y$  and thus to  $(2y)(\psi y \supset \chi y)$ . But where the  $f$

itself is a variable, as it is in formula  $F$ , nothing of this sort is done. We could indeed (following Stanisław Leśniewski) introduce a symbol for the predicate “ $\chi$ -ing-if-one- $\phi$ ’s” by a special definition; for example

$$[\supset \psi \chi]x =_{\text{Df}} \phi x \supset \psi x$$

and so replace  $F$  with  $G$ :

$$(f) : (\phi)f(\phi) . \supset .f([\supset \psi \chi]),$$

but then it would be impossible to eliminate the defined symbol from  $G$  in favor of the symbols by which it is defined, and it seems an odd sort of definition that would be thus limited. (Church’s use of  $\lambda$  can in fact be regarded as simply a generalization of Leśniewski’s procedure.)

The uneliminability of “abstracts” from these contexts is an odd and perhaps awkward fact, but it need not be taken to imply that there are abstract objects, for “abstracts” need not be regarded as a kind of name. In expositions of the  $\lambda$ -calculus it is often said that the form  $\lambda x \phi x$  corresponds to the ordinary-language quasi noun “ $\phi$ -ing,” but this is not strictly correct, as may be seen from the fundamental equation  $(\lambda x \phi x)a = \phi a$ . If “ $\phi$ ” here represents not a name but a verb (“ $\phi a$ ” means “ $a$   $\phi$ ’s”), then so must “ $\lambda x \phi x$ ” (“ $(\lambda x \phi x)a$ ” also means “ $a$   $\phi$ ’s”), so that if  $f$  in  $f(\phi)$  is a function with not names but predicates as arguments, so it must be in “ $f(\lambda x \phi x)$ .”

**RAMIFIED THEORY OF TYPES.** We may now describe the added feature that makes Russell’s own presentation of his theory of types more complex than the presentation so far given here. Russell divides functions into types not only according to the types of argument that they take but also according to whether they do or do not involve an internal reference to all functions of (what appear to be) their own type. For example, the function  $\hat{x}$  has all the qualities of a great general” has individual-name arguments, just as “ $\hat{x}$  is brave” does, but unlike “ $\hat{x}$  is brave” it has a “for all  $\phi$ ” within itself—it amounts to “For all  $\phi$ , if whoever is a great general  $\phi$ ’s, then  $\hat{x}$   $\phi$ ’s.” Russell therefore regards it as of a different type, or, as he often says, of a different order, from “ $\hat{x}$  is brave.” Functions that do not thus involve a reference to all functions of (what appear to be) their own type he calls “predicative” functions and symbolizes them by putting an exclamation mark or “shriek” after the symbol, as in “ $\phi!x$ .” Functions cannot in fact (on Russell’s view) strictly contain references to all functions of *their own* type or order but references only to ones of orders below their own. A function of individuals, which contains a reference to all predicative functions of individuals, is not itself predica-

tive and cannot be regarded as being among the functions to which it implicitly refers. Having all the properties of a great general, for example, is not itself a property of a great general, at least not in the same sense of “property”—it is a second-order property.

What this means in practice might be illustrated as follows: It seems that if there were no facts about  $x$ —that is, if for no  $\phi$ ,  $\phi x$ —then there would be at least one fact about  $x$ , namely the fact that there are no facts about it, and hence it cannot be that there are no facts about  $x$ . In symbols, from

$$(1) \quad \psi x \supset (\exists \phi)(\phi x)$$

it seems possible to obtain

$$(2) \quad \sim(\exists \phi)(\phi x) . \supset (\exists \phi)(\phi x)$$

by letting  $\psi \hat{x}$  in (1) be, in particular,  $\sim(\exists \phi)(\phi \hat{x})$ ; and from (2) it follows by a kind of *reductio ad absurdum* that for any given  $x$  we have  $(\exists \phi)(\phi x)$ . But on Russell’s view this proof will not do, for (1) ought to have been written

$$(3) \quad \psi!x \supset (\exists \phi)(\phi!x)$$

and here  $\sim(\exists \phi)(\phi!x)$ , not being itself predicative, is not a permissible substitution for  $\psi!x$ . It is worth noting, however, that our final conclusion,  $(\exists \phi)(\phi!x)$ , can be proved from (3) in a different way—by letting our  $\psi!x$  be  $\chi!x \supset \chi!x$  (“ $\hat{x}$   $\chi$ ’s-if-it- $\chi$ ’s”), which *is* predicative and is true of any  $x$ , so that what it implies must be true of any  $x$  also. (The new argument is as follows: There is always some fact about  $x$ , since at least it is a fact that  $x$  is red-if-it-is-red, square-if-it-is-square, etc.)

**Axiom of reducibility.** Russell lumps together all his type and order restrictions under the general head of avoiding “vicious circles,” and the theory of types with the theory of orders worked into it is called the “ramified” theory of types. One trouble with it is that it vitiates certain essential arguments in the higher reaches of mathematics, and to save these Russell introduced an “axiom of reducibility,” that to every function of any order there corresponds a predicative function that is formally equivalent to it—that is, which holds for exactly the same arguments as the given function. This means that any argument like our allegedly invalid proof of  $(\exists \phi)(\phi x)$  above, where it is worth saving, can in principle be replaced by one like our second and valid one; the axiom of reducibility does not itself enable us to find this valid argument but entitles us to proceed as if we had it. It is, however, an intuitively dubious principle and can be dispensed with if we can content ourselves with the theory

of types in the “simple” form in which it has been stated in earlier sections.

**Semantic paradoxes.** It was pointed out by F. P. Ramsey that those paradoxes which Russell lists and which cannot be eliminated (as can, for example, the paradox of the class of all classes not members of themselves) by the “simple” theory of types always contain some implicitly or explicitly “semantic” feature; that is, they all have to do with the relation of language to what it is about and all involve conceptions like truth and meaning. A typical example is the paradox of the liar, of the man who says “What I am now saying is false” and says nothing else but this, so that what he says is true if it is false and false if it is true. Such paradoxes are now generally dealt with by assuming not only a hierarchy of “parts of speech” in one’s basic language (this is what the simple theory of types amounts to) but also a hierarchy of languages—a basic language, a “metalanguage” in which we discuss the meaning and truth of expressions in the basic language, a “metametalinguage” in which we deal similarly with the metalanguage, and so on.

It is both easy and necessary to criticize Russell’s theories concerning the logical and semantic paradoxes, and his work in logic and the foundations of mathematics generally, but he remains, more than any other one person, the founder of modern logic.

## ETHICS AND THE CRITIQUE OF RELIGION

**ETHICS.** Much of Russell’s life, as we saw in an earlier section, was devoted to the advocacy of certain moral and political ideals. In this sense of the word *moralist*, in which it has no derogatory implications, Russell was certainly a moralist and frequently a very passionate one at that. Unlike many other moralists he was also concerned with what are now referred to as “metamoral” or “metaethical” issues. He repeatedly addressed himself to questions about the status of moral principles—what, if anything, they mean, what kind of disagreement there is between people who support opposite moral positions, and whether inferences from nonmoral premises to a moral conclusion can ever be valid. In discussing Russell’s ethics, we will be concerned only with his metamoral theories.

**Early views.** In his first important essay on this subject, “The Elements of Ethics” (1910), Russell defended a position closely akin to that of G. E. Moore in *Principia Ethica*. “Good and bad,” he wrote, “are qualities which belong to objects independently of our opinions, just as

much as *round* and *square* do; and when two people differ as to whether a thing is good, only one of them can be right, though it may be very hard to know which is right.” The goodness or badness of a thing cannot be inferred from *any* of its other properties. “Knowledge as to what things exist, have existed, or will exist, can throw absolutely no light upon the question as to what things are good.” Russell was by no means unaware at this time of the wide appeal of the familiar arguments for subjectivism—the “divergence of opinion” on moral questions and the difficulty of “finding arguments to persuade people who differ from us in such a question” (“The Elements of Ethics,” in *Readings in Ethical Theory*, edited by Wilfrid Sellars and John Hospers, New York, 1952, pp. 6–7). But he did not then regard these arguments as having any logical force. “Difficulty in discovering the truth,” he wrote, “does not prove that there is no truth to be discovered” (p. 6). Like Moore, he argued that if subjectivism were true it would follow that in a moral dispute there is never really any “difference of opinion” between the disputing parties. If when *A* says *x* is good and *B* says *x* is bad, *A* and *B* were really talking about their respective feelings or desires, they might well both be right at the same time and “there would be no subject of debate between them.” At that time Russell regarded this as plainly false. “As a matter of fact,” he observed, “we consider some tastes better than others: we do not hold merely that some tastes are ours and other tastes are other people’s.” When “The Elements of Ethics” was reprinted in 1952 in *Readings in Ethical Theory*, the anthology mentioned above, Russell added a footnote in which he explained that “not long after publishing this paper [he] came to disagree with the theory that it advocates.” He explains that the change in his views was originally due to George Santayana’s criticisms in his *Winds of Doctrine*, but he adds that he “found confirmation” for his later position “in many other directions.” Russell’s later position was first mentioned very briefly in a 1921 preface to a paperback reprint of “A Free Man’s Worship”; it was explained in some detail in *What I Believe* (1925) and in *The Outline of Philosophy* (1927), and it received its fullest formulations in *Religion and Science* (1935), *Power* (1938), “Reply to My Critics” (in P. A. Schilpp, ed., *The Philosophy of Bertrand Russell*, 1944), and *Human Society in Ethics and Politics* (1955).

**The subjectivity of values.** Except on one basic issue, Russell’s later position is a point-by-point denial of the earlier theory. “Good” and “bad” are no longer regarded as qualities belonging to objects, and in this respect they are now explicitly contrasted with “square” and “sweet”: “If two men differ about values, there is not a disagree-

ment as to any kind of truth, but a difference of taste” (*Religion and Science*, pp. 237–238); “There are no facts of ethics” (*Power*, p. 257); “I see no property analogous to truth that belongs or does not belong to an ethical judgment” (“Reply to My Critics,” p. 723). “Taste” in the first of these passages is used in a very broad sense to cover all kinds of psychological states and attitudes, including desires. Russell does not, of course, deny the plain fact that people regard some tastes as better than others and some desires as higher than other desires, but now he is willing to maintain that this merely means that the tastes or desires are their own. “What we ‘ought’ to desire is merely what someone else wishes us to desire” (*What I Believe*, p. 29).

Russell is quite ready to have his later theory classified as a form of “the doctrine of the subjectivity of values” (*Religion and Science*, p. 237), but it differs in some significant respects from the older theories that have gone by that name. If somebody maintains that pleasure, for example, or the love of God, is intrinsically good, or good “on its own account,” this must not be taken to be equivalent to the statement that he approves of it or in some way desires it. Like the advocates of the so-called emotive theory of ethics, Russell maintains that intrinsic moral judgments, grammatical appearances notwithstanding, are not statements or assertions at all but *expressions* of desire. “A judgment of intrinsic value,” he writes in *Power*, “is to be interpreted, not as an assertion, but as an expression of desire concerning the desires of mankind. When I say ‘hatred is bad,’ I am really saying: ‘would that no one felt hatred.’ I make no assertion; I merely express a certain type of wish” (*Power*, p. 257).

Both here and in his capacity as a reformer Russell places much emphasis on the distinction between purely personal and what he calls “impersonal” desires. A hungry man’s desire for food or an ambitious man’s desire for fame are examples of the former; a desire for the abolition of the death penalty or the end of racial discrimination, independently of whether the person in question stands to gain from these changes, are examples of the latter. In moral judgments we express certain of our impersonal desires. A king who says, “Monarchy is better than republican forms of government,” is using the word *better* in its properly moral sense if he is expressing not just his desire to remain king but a desire that nations have monarchical systems regardless of his own personal position. Russell occasionally writes as if the desire expressed by moral judgments must be a second-order desire—that is, a desire that everybody have a certain first-order desire—but as several of his own examples make clear,

this is not part of his position. What is essential is that the desire be impersonal. In this connection he also observes that the philosophers who stressed the “universality” of moral principles were in a sense quite right. This universality, however, does not consist in any a priori character or logical necessity. What is universal is the *object of the desire* expressed by a moral judgment. “The wish, as an occurrence, is personal, but what it desires is universal... It is this curious interlocking of the particular and the universal which has caused so much confusion in ethics” (*Religion and Science*, p. 236).

As we shall see, Russell had a tendency to overestimate the scope of application of his subjectivism, but in a number of places he points out quite explicitly that large classes of everyday moral judgments and disputes do not come within the purview of the theory. “Ethical controversies are very often as to means, not ends” (*Power*, p. 259). “The framing of moral rules, so long as the ultimate Good is supposed known, is matter for science” (*Religion and Science*, p. 228). It follows from this that if human beings could agree about ultimate ends, all moral disputes would in principle be decidable by an appeal to facts even though the intrinsic judgments would still be not bona fide propositions but expressions of wishes. In fact, however, Russell insists, there is no such agreement about ends. In “The Elements of Ethics” he had conceded that there were *some* ultimate ethical differences but had maintained that people in fact “differ very little in their judgments of intrinsic value.” Many of the commonly observed differences are wrongly regarded as ultimate because what are really disagreements about means are mistaken for disagreements about ends. In his subjectivist phase Russell seems to think that differences about ends are not at all uncommon. Behind such disputes as, for example, the subjection of women or the persecution of religious minorities, which do involve questions of means, he writes, “there is generally a difference as to ends,” and this sometimes becomes “nakedly apparent,” as in Friedrich Nietzsche’s criticisms of Christian ethics. In Christianity, all men are valued equally, but for Nietzsche the majority exists only as means to the superman. This, Russell maintains, is an example of a dispute about ends, and “it cannot be conducted, like scientific controversies, by appeals to facts” (*Power*, p. 259).

In “The Elements of Ethics” Russell had quite properly observed that the mere existence of widespread ethical disagreement (if it is indeed widespread) does not establish any form of subjectivism. Although he has evidently come to believe that ethical disagreement is more widespread than he had thought earlier, he does not offer



this as evidence for his new theory. What he does offer as evidence is the undecidability of ethical disputes. He writes:

[The chief ground for adopting this view] is the complete impossibility of finding any arguments to prove that this or that has intrinsic value.... We cannot *prove*, to a color-blind man, that grass is green and not red. But there are various ways of proving to him that he lacks a power of discrimination which most men possess, whereas in the case of values there are no such ways ... since no way can be even imagined for deciding a difference as to values, the conclusion is forced upon us that the difference is one of taste, not one as to any objective truth. (*Religion and Science*, p. 238)

If three men argue, one saying “The good is pleasure,” the second “The good is pleasure for Aryans and pain for Jews,” and the third “The good is to praise God and glorify him forever,” they cannot, as people engaged in a scientific dispute, “appeal to facts,” for facts, it seems obvious, “are not relevant to the dispute” (*Power*, p. 257).

Russell’s later view agrees with the earlier position on only one significant point, its opposition to naturalism. By “naturalism” is here meant the theory that there *is* a logical connection between some moral judgments and factual premises where the latter are not necessarily confined to empirical statements but may also include metaphysical doctrines. We saw how in “The Elements of Ethics” Russell had insisted that from statements concerning what exists nothing can be inferred about “the goodness of anything.” “It is logically impossible,” he repeated in the course of expounding his later position, “that there should be evidence for or against” a moral judgment, but now this is maintained because a moral judgment “makes no assertion” and hence possesses neither truth nor falsehood (*Religion and Science*, pp. 236–237).

**“Incredibility” of Russell’s subjectivism.** Rather than attempt a detailed critical evaluation of Russell’s subjectivism, we will discuss one objection that has been urged by a number of his critics and which, in one form or another, has been leveled against nearly all forms of subjectivism. It has been argued that a subjectivist cannot consistently make moral judgments. All he can say is that some people have one kind of feeling or attitude while other people feel differently. More specifically, how can Russell’s subjectivism be reconciled with his judgments as a moral critic and reformer?

It may be replied that as a matter of pure logic there is no inconsistency between holding that moral judgments are expressions of taste and using moral language to express one’s own tastes. Russell, it might be said, would be inconsistent only if he claimed that *his* moral judgments, unlike those of his opponents, are more than expressions of taste. Then he would indeed be like the man who, in the course of an argument about the value of a piece of music, remarked to his opponent “It is all a matter of taste, except that my taste is better than yours.” However, while this answer is valid as far as it goes, it does not meet the heart of the objection. For Russell seems to be saying—or at least he would like to be able to say—that his moral judgments (for example, his judgment that democracy is a better system than totalitarianism or that the sexual code advocated in *Marriage and Morals* is superior to that associated with orthodox religion) are in some sense rational or right or well-grounded while the judgments of his opponents are irrational, wrong, or unsupported by the evidence.

Russell apparently did not, when he first advanced his subjectivism, see any serious problem here, but in the 1940s and 1950s he repeatedly expressed dissatisfaction with his own theory on this ground. Thus, in “Reply to My Critics” he writes:

What are “good” desires? Are they anything more than desires that you share? Certainly there *seems* to be something more. Suppose, for example, that some one were to advocate the introduction of bull-fighting in this country. In opposing the proposal, I should *feel*, not only that I was expressing my desires, but that my desires in the matter are *right*, whatever that may mean. As a matter of argument, I can, I think, show that I am not guilty of logical inconsistency in holding to the above interpretation of ethics and at the same time expressing strong ethical preferences. But in feeling I am not satisfied. (*The Philosophy of Bertrand Russell*, p. 724)

To this he adds: “I can only say that, while my own opinions as to ethics do not satisfy me, other people’s satisfy me still less.” More than a decade later Russell expressed himself even more strongly. In a letter to the *Observer* (October 6, 1957) he comments on Philip Toynbee’s review of *Why I Am Not a Christian*: “What Mr. Toynbee says in criticism of my views on ethics has my entire sympathy. I find my own views argumentatively irrefutable, but nevertheless incredible. I do not know the solution.”

It is doubtful whether in such comments Russell is really fair to his own subjectivism. Let us recall that the

theory was never meant to apply to anything other than what are variously called intrinsic or fundamental value judgments and differences. The questions whether happiness is better than unhappiness and love better than hate are frequently cited as such ultimate moral issues, but it would be hard to find anybody who seriously maintains that suffering is good on its own account or that hate is better than love, although of course people have often maintained that in certain situations and for certain reasons suffering and hate are preferable to enjoyment and love. However, on occasions there do appear to be real value differences of an ultimate kind. Thus, some people would maintain that dignity is “more important” or “nobler” than happiness. Many who do not despise happiness at all would maintain without hesitation that a man who chose to suffer a great deal rather than compromise his integrity (where it is assumed that he would in fact have suffered much less if he had not stood his ground) lived a better life than he would have if he had made the opposite choice. Or, again, there is sometimes disagreement as to whether a person suffering from a fatal illness should be told the truth, although there may be full agreement about the consequences of both telling and not telling him the truth. Russell’s subjectivism does apply to this kind of intrinsic moral disagreement, and in such situations he could not, consistently with his theory, claim that the moral judgment he endorses is “more rational” or better supported than that of his opponents.

However, the examples Russell offers when expressing dissatisfaction with his subjectivism are not at all of this ultimate kind, and this applies to all or nearly all the positions he has advocated in his social and political writings. The man who says that the good is pleasure for Aryans and pain for Jews, if he is willing to engage in moral argument at all—if he is not, the problem does not arise—presumably does not *just* say this but proceeds to make all kinds of factual claims about the psychological and physical qualities of Aryans and Jews, respectively, about the laws of heredity, and about various other matters that he regards as justifying his moral position. Similarly, the man who maintains that “the good is to praise God and glorify him forever” presupposes that there is a God, and a God of a certain kind, probably also that he has revealed himself in certain ways, and, if challenged (or perhaps even without being challenged), he will make claims about the hollowness of all earthly satisfactions and the greater reliability, intensity, and duration of the satisfactions derived from glorifying God. Again, a man, who advocates the introduction of bullfighting into the United States would not *just* advance this proposal but would give reasons having to do, perhaps, with the bene-

fits to be derived from engaging in dangerous sports and the special thrills experienced by the spectators. All these supporting factual claims are discussable, and it may be possible to show that they are mistaken or highly implausible. If so, it might well be possible to regard the case of one side in such a dispute as well supported and the other as unsupported by the evidence. In all cases in which the person is willing to support his moral judgment by factual premises, it is perfectly consistent for Russell to assert that one position is “more rational” than the other, where “more rational” does not merely mean that Russell shares the attitude of the person taking this position.

What seems to be amiss here is not Russell’s subjectivism but his view (which is not logically implied by it) that the theory applies to cases like the dispute about bullfighting. In his later period Russell seems to be guilty of a gross overestimate of the prevalence of ultimate moral disagreements. It is true, as he observes in *Power*, that behind disagreements about means there is frequently disagreement about ends, but it is very doubtful that the ends in question are in most cases *ultimate* ends. To give a simple illustration of a very common type: Two people may offer conflicting moral judgments about a bill to legalize abortion. The man who opposes the legislation may give as his reason (or as one of his reasons) that it would remove one of the conditions restraining unmarried people from engaging in sexual intercourse, whereas the other man might offer this as his reason (or one of his reasons) for supporting the legislation. Although the disagreement may in the immediate context be properly described as one about an end, it is clearly not about an ultimate end. In all likelihood the parties to the dispute would differ about the effects of a freer sex life on personal happiness, on society at large, on the future of religious institutions, and many other things. It is doubtful that either of them would maintain that suffering as such is better than happiness or that hate is better than love.

Even people who advocate what by most contemporary standards would be regarded as “outlandish” moral positions can usually be seen to share many of the intrinsic value judgments of the rest of humankind. Thus, Arthur Schopenhauer and other champions of asceticism recommend the suppression of desires, including those that to most human beings seem the most natural and the most innocent, but they do so *not* because in their opinion suppression of these desires would make people unhappy but, on the contrary, because it would enable them to achieve greater happiness or at least because it would reduce suffering to a minimum. In Norman Mailer’s bizarre novel *An American Dream* the main char-

acter offers a defense of murder, but this unusual position is justified by the argument that “murder offers the promise of vast relief. It is never unsexual.” It is accompanied by “exhilaration” that must come “from possessing such strength.” It should be noted that murder is here justified not because it causes suffering but because, according to the character, it leads to “exhilaration.” In other writings Mailer tells us that the “modern soul marooned in ... emptiness, boredom and a flat, dull terror of death” would be well advised to pass through “violence, cannibalism, insanity, perversion” and other states and activities that are usually considered highly undesirable, but these recommendations are offered not for their own sake but because they will lead the person “back to life.”

As for the really intrinsic clashes of the kind mentioned earlier, to which Russell’s subjectivism would apply, one wonders if the consequences of the theory are there really so paradoxical. No doubt people do in such disputes regard their position as superior to that of their opponents—the man who admires integrity will feel contempt for the “cowardly” compromiser, and the compromiser will think the man who chooses to suffer a fool. Here, however, unless there are some *hidden* differences concerning matters of fact, it seems not at all incredible to maintain that calling one position superior simply amounts to expressing one’s own preference for it.

None of the above is meant to prove that Russell’s subjectivism is a correct account of the logical status of moral judgments, but it would indicate that the favorite objection of his critics can be disposed of without much difficulty.

**CRITIQUE OF RELIGION.** No such doubts as Russell has expressed about his subjectivism in ethics mark his views on religion. Unlike many academic philosophers whose position is very similar to his, Russell did not hesitate to express his convictions publicly and without equivocation or compromise. Ever since he abandoned the Platonic theory of ideas, Russell was a forthright opponent of religion in more senses than one: He regards the basic doctrines of (supernaturalistic) religions as intellectually indefensible, he argues that religious belief has not on balance been a force for good but quite the opposite, and he hopes and believes that religion will eventually die out. “I am myself,” he wrote in 1922, “a dissenter from all known religions, and I hope that every kind of religious belief will die out.... I regard religion as belonging to the infancy of human reason and to a stage of development which we are now outgrowing” (*Sceptical Essays*, p. 101). In a television interview thirty-seven years later he

slightly qualified this prediction. If great wars and great oppressions continue so that many people will be leading very unhappy lives, religion will probably go on, but “if people solve their social problems religion will die out” (*Bertrand Russell Speaks His Mind*, p. 31).

**God.** Russell wavered between calling himself an agnostic and describing himself as an atheist. He evidently did not attach too much importance to this distinction, but he had made it clear that if he is to be classified as an agnostic, it would have to be in a sense in which an agnostic and an atheist are “for practical purposes, at one.” In the television interview mentioned earlier the interviewer asked Russell, “Do you think it is certain that there is no such thing as God, or simply that it is just not proved?” “No,” Russell answered, “I don’t think it is certain that there is no such thing—I think that it is on exactly the same level as the Olympic gods, or the Norwegian gods; they also may exist, the gods of Olympus and Valhalla. I can’t prove they don’t, but I think the Christian God has no more likelihood than they had. I think they are a bare possibility” (*Bertrand Russell Speaks His Mind*, pp. 24–25). He explained his views more fully in an interview published in *Look* magazine in 1953. An agnostic, in any sense in which he can be regarded as one, Russell said, “may hold that the existence of God, though not impossible, is very improbable; he may even hold it so improbable that it is not worth considering in practice” (Leo Rosten, ed., *A Guide to the Religions of America*, New York, 1955, p. 150).

**Immortality.** On survival, Russell’s position is similarly negative. All the evidence indicates that what we regard as our mental life is “bound up with brain structure and organized bodily energy.” There is every reason to believe that mental life ceases when the body decays. Russell admits that this argument is “only one of probability” but adds that “it is as strong as those upon which most scientific conclusions are based” (*Why I Am Not a Christian*, p. 51). It is conceivable that evidence from psychological research might change the balance of probability some day, but, writing in 1925, Russell considered such evidence far weaker “than the physiological evidence on the other side.” He did not later see any reason to modify this judgment.

Russell’s views on the body-mind problem are known as “neutral monism,” and it would be inaccurate to call him a materialist. However, he always emphasized that as a theory about man’s place in the universe his philosophy is closely akin to materialism. “Emotionally,” he wrote in 1928, “the world is pretty much the same as it would be if the materialists were in the right” (*In Praise of*

*Idleness*, p. 143). The opponents of materialism, he adds, have been actuated by the desire to prove that the mind is immortal and that the “ultimate power” in the universe is mental and not physical. On both these points, Russell makes clear, he agrees with materialism. When he returned to the subject in 1959 he had not changed his opinion at all. “I still think,” he wrote then, “that man is cosmically unimportant, and that a Being, if there were one, who could view the universe impartially, without the bias of *here* and *now*, would hardly mention man, except perhaps in a footnote at the end of the volume” (*My Philosophical Development*, p. 213).

**Objections to fideism.** Although, needless to say, Russell rejected the traditional arguments for the existence of God and immortality, he greatly preferred the rationalistic theology of such philosophers as Thomas Aquinas and Descartes to the fideism of Blaise Pascal, Jean-Jacques Rousseau, Søren Kierkegaard, and their numerous modern followers. “The rejection of reason in favor of the heart,” he writes, “was not, to my mind, an advance.” He remarks that “no one thought of this device so long as reason appeared to be on the side of religious belief” (*A History of Western Philosophy*, p. 720). There are two fatal objections to the practice of justifying religious belief by an appeal to the emotions of the heart. To begin with, the heart says different things to different men and to the same man at different times, but even if the heart said the same thing to all men this would still not be evidence for the existence of anything outside our emotions, and the fideists, no less than the rationalistic believers, mean to make claims about objective fact, not merely about their own emotions. At bottom, Russell concludes, the only reason offered for the acceptance of the new theology is “that it allows us to indulge in pleasant dreams. This is an unworthy reason, and if I had to choose between Thomas Aquinas and Rousseau, I should unhesitatingly choose the Saint” (*My Philosophical Development*, p. 721).

Some unbelievers have gone out of their way to praise the greatness of Jesus and to admit that religious belief, although perhaps not true, is at least of great value to individual believers and to society. Russell makes no such concessions. Although he grants that some of Christ’s maxims were indeed admirable (especially those consistently disregarded by Christian dignitaries) he finds much in the teachings of Jesus to be defective, in particular his doctrine of eternal damnation. “Either in the matter of virtue or in the matter of wisdom,” Russell concludes, Christ does not “stand as high as some other people known to history”—for example, Buddha and Socrates (*Why I Am Not a Christian*, p. 19).

**Harmfulness of religious belief.** Russell’s views about the nature of the emotions that inspire religious belief (“it is based, primarily and mainly, upon fear”) and also about the harmful influence of religious organizations are very similar to those of David Hume, Baron d’Holbach, and other eighteenth-century freethinkers. He did, however, devote rather more attention to the bad effects of the habit of accepting propositions on faith—in the absence of or even in opposition to the evidence. It is an error, Russell contends, to suppose that a person who does not form his beliefs on the basis of evidence in one domain can remain open-minded and scientific in another. Furthermore, somebody holding comfortable beliefs on faith dimly realizes that they are myths and “becomes furious when they are disputed.” Such a person will therefore do his best to suppress all critics who might remind him of the feeble backing of his beliefs. Russell makes it clear that in this context he is not criticizing Christianity only. “The important thing,” he writes, “is not what you believe, but how you believe it.” The objections to “faith” do not depend on what the faith in question may be. “You may believe in the verbal inspiration of the Bible or of the Koran or of Marx’s *Capital*. Whichever of these beliefs you entertain, you have to close your mind against evidence; and if you close your mind against evidence in one respect, you will also do so in another, if the temptation is strong.” The person who bases his belief on reason will support it by argument rather than by persecution and will abandon his position if the argument goes against him. If, however, his belief is based on faith, he will conclude that argument is useless and will “therefore resort to force either in the form of persecution or by stunting and distorting the minds of the young whenever he has the power to control their education” (*Human Society in Ethics and Politics*, pp. 207–208).

**“The world is horrible.”** Russell never denied that in some respects a “godless” philosophy like his has to be gloomy. The beginning of wisdom, he teaches, is acceptance of the fact that the universe does not care about our aspirations and that happiness and unhappiness are not meted out in accordance with what people deserve. “The secret of happiness,” he observed during a television program commemorating his ninety-second birthday, “is to face the fact that the world is horrible.” What Russell meant by this becomes clear from a story related by his biographer, Alan Wood. Wood’s wife had expressed her opinion that it seemed horribly unjust that the young men who had been killed in the war should not somehow or somewhere have a second chance to achieve happiness. “But the universe *is* unjust,” Russell replied, “the secret of happiness is to face the fact that the world is horrible,

horrible, *horrible* ... you must feel it deeply and not brush it aside ... you must feel it right here”—hitting his breast—“and then you can start being happy again” (*Bertrand Russell: The Passionate Sceptic*, p. 237). Once a person has stopped looking at the universe in terms of anthropomorphic demands, he can concentrate on what is attainable and not waste his time in self-pity and cosmic complaints. For those whose philosophy is shaped not by a respect for facts but by their wishes Russell was always scathing in his contempt. He expressed his amazement that courage is praised in all types of situations but not when it comes to forming a view about the world. “Where traditional beliefs about the universe are concerned,” he writes, “craven fears ... are considered praiseworthy, while intellectual courage, unlike courage in battle, is regarded as unfeeling and materialistic.” Writing in 1957, he notes that this attitude is perhaps less widespread than it was in his youth, but he adds that it “still inspires vast systems of thought which have their root in unworthy fears.” “I cannot believe,” he concludes, that there can ever be any good excuse for refusing to face the evidence in favor of something unwelcome. It is not by delusion, however exalted, that mankind can prosper, but only by unswerving courage in the pursuit of truth” (*Fact and Fiction*, p. 46).

**See also** Absolute, The; Asceticism; Analysis, Philosophical; Balfour, Arthur James; Bradley, Francis Herbert; Church, Alonzo; Correspondence Theory of Truth; Descartes, René; Epistemology, History of; Ethical Subjectivism; Existence; Frege, Gottlob; Hegel, Georg Wilhelm Friedrich; Hegelianism; Holbach, Paul-Henri Thiry, Baron d'; Hume, David; Infinity in Mathematics and Logic; James, William; Kierkegaard, Søren Aabye; Logical Paradoxes; Logic, History of; Logic, Modern; Logic, Traditional; Luther, Martin; Mathematics, Foundations of; McTaggart, John McTaggart Ellis; Memory; Metaethics; Mill, John Stuart; Mind-Body Problem; Modal Logic; Moore, George Edward; Neumann, John von; Nietzsche, Friedrich; Number; Pascal, Blaise; Peano, Giuseppe; Peirce, Charles Sanders; Plato; Platonism and the Platonic Tradition; Pluralism; Proper Names and Descriptions; Propositions; Quantifiers; Quine, Willard Van Orman; Ramsey, Frank Plumpton; Realism; Rousseau, Jean-Jacques; Santayana, George; Schopenhauer, Arthur; Sellars, Wilfrid; Socrates; Thomas Aquinas, St.; Types, Theory of; Voltaire, François-Marie Arouet de; Whitehead, Alfred North; Wittgenstein, Ludwig Josef Johann.

## Bibliography

### BIOGRAPHY

There is a good deal of autobiographical material in Russell's *Portraits from Memory and Other Essays* (New York: Simon and Schuster, 1956); in *Fact and Fiction* (New York: Simon and Schuster, 1962); in his introduction to *Selected Papers of Bertrand Russell* (New York: Modern Library, 1927); in “My Religious Reminiscences,” in *Rationalist Annual* 55 (1938): 3–8; in “My Mental Development,” in *The Philosophy of Bertrand Russell*, edited by P. A. Schilpp (Evanston and Chicago: Open Court, 1944); and in *My Philosophical Development* (New York: Simon and Schuster, 1959). Alan Wood, *Bertrand Russell: The Passionate Sceptic* (London: Allen and Unwin, 1957), is the only full-length biographical study of Russell. H. W. Leggett, *Bertrand Russell* (New York, 1950), is a short pictorial biography.

- G. H. Hardy, *Bertrand Russell and Trinity* (Cambridge, U.K.: Cambridge University Press, 1942), traces the controversy between Russell and the fellows of Trinity College over his pacifist activities during World War I. *Rex versus Bertrand Russell, Report of the Proceedings before the Lord Mayor* (London, 1916), gives the text of the first of Russell's trials.
- D. H. Lawrence, *Letters to Bertrand Russell* (New York: Gotham Book Mart, 1948), reproduces Lawrence's letters to Russell during World War I; Russell's letters to Lawrence have not been preserved.

Russell's part in the Beacon Hill School is most fully described in Joe Park, *Bertrand Russell on Education* (Columbus: Ohio State University Press, 1963). The Park volume also contains a complete list of Russell's writings on educational topics. Details about the City College case of 1940 can be found in *The Bertrand Russell Case*, edited by John Dewey and Horace M. Kallen (New York: Viking Press, 1941); in a publication by the American Civil Liberties Union titled *The Story of the Bertrand Russell Case—The Enlightening Record of the Obstruction by Courts and Officials of the Appointment of Bertrand Russell to a Professorship at the College of the City of New York* (New York: American Civil Liberties Union, 1941); and in Paul Edwards, “How Bertrand Russell Was Prevented from Teaching at City College,” which is an appendix to Russell's *Why I Am Not a Christian and Other Essays on Religion and Related Subjects* (London: Allen and Unwin, and New York: Simon and Schuster, 1957).

### EPISTEMOLOGY AND METAPHYSICS

*Principles of Mathematics* (Cambridge, U.K.: Cambridge University Press, 1903) was Russell's first major philosophical work. Its position is one of Platonic realism. In the preface to the second edition (1937) Russell sets forth his later disenchantment with this position. For a nonmathematical exposition of Russell's early realism, see “Meinong's Theory of Complexes and Assumptions,” in *Mind* 13 (1904): 204–219; 336–354; 509–524. Russell's criticisms of the idealist theory of truth are to be found in “The Monistic Theory of Truth,” in *Philosophical Essays* (New York: Longman, 1910), a revised version of “The Nature of Truth,” in *Mind* 15 (1906): 528–533. *Philosophical Essays* also contains two influential essays by Russell attacking the pragmatist theory of truth.

The shift from realism to logical constructionism can be followed in a number of articles, the most important of

which is "On Denoting," in *Mind* 14 (1905): 479–493. This, together with other important but otherwise largely unavailable essays, is reprinted in Russell's *Logic and Knowledge*, edited by R. C. Marsh (London: Allen and Unwin, 1956). Russell's "On the Relations of Universals and Particulars," in *PAS* 12 (1911–1912): 1–24, reprinted in *Logic and Knowledge*, is a classic presentation of the largely Platonic theory of universals Russell still held at that time. *Problems of Philosophy* (New York: Holt, 1912) gives an excellent semipopular account of the general state of Russell's thinking then. Russell's early attempts to represent physical objects as logical constructions can be seen in *Our Knowledge of the External World* (Chicago: Open Court, 1914) and in two essays, "The Ultimate Constituents of Matter," in *Monist* 25 (1915): 399–417, and "The Relations of Sense-Data to Physics," in *Scientia* (4) (1914), both reprinted in *Mysticism and Logic* (London: Allen and Unwin, 1918). Other important essays in this collection are "On Scientific Method in Philosophy" (1914); "On the Notion of Cause," originally published in *PAS* 13 (1912–1913): 1–26; and "Knowledge by Acquaintance and Knowledge by Description," originally published in *PAS* 11 (1910–1911): 108–128. See also "The Philosophy of Logical Atomism," in *Monist* 28 (1918): 495–527; 29 (1919): 32–63, 190–222, and 345–380; reprinted in *Logic and Knowledge* (see above). The analysis of basic concepts and principles of physical science is pushed further in *The Analysis of Matter* (New York: Harcourt Brace, 1927). Logical constructionism is applied to mental phenomena in *The Analysis of Mind* (New York: Macmillan, 1921). Russell's increasing concern with psychological aspects of meaning can be traced in "On Propositions, What They Are and How They Mean," in *PAS*, supp. 2 (1919): 1–43, reprinted in *Logic and Knowledge*, in Ch. 10 of *The Analysis of Mind*; and in Russell's most extensive work on meaning and empirical data, the rich but chaotic *An Inquiry into Meaning and Truth* (New York: Norton, 1940). Russell's later thoughts on meaning and various other problems concerning empirical knowledge, particularly in the physical sciences, are given a relatively systematic presentation in *Human Knowledge, Its Scope and Limits* (New York: Simon and Schuster, 1948).

In several works Russell summarized his philosophy and/or its development. The most important of these are "Logical Atomism," in *Contemporary British Philosophy*, edited by J. H. Muirhead, first series (London: Allen and Unwin, 1924), reprinted in *Logic and Knowledge* (see above); "My Mental Development," in *The Philosophy of Bertrand Russell*, edited by P. A. Schilpp (see above); and the very interesting recent work *My Philosophical Development* (New York: Simon and Schuster, 1959). The last-named work also contains some of Russell's polemics against Oxford philosophers and their criticisms of his views. Russell's *A History of Western Philosophy* (New York: Simon and Schuster, 1946) and *The Wisdom of the West* (New York: Doubleday, 1959), aside from their intrinsic interest, are of great value to students of Russell's thought in showing us his mature evaluations of the great philosophers of past ages.

The critical literature on different aspects of Russell's epistemology and metaphysics is vast. *The Philosophy of Bertrand Russell* (see above) contains a number of excellent discussions, together with Russell's replies. Special mention should also be made of C. A. Fritz, *Bertrand Russell's*

*Construction of the External World* (London: Routledge & K. Paul, 1952); Erik Götlind, *Bertrand Russell's Theories of Causation* (Uppsala: Almqvist and Wiksells, 1952); J. O. Urmson, *Philosophical Analysis: Its Development between Two World Wars* (Oxford: Clarendon Press, 1956); and G. J. Warnock, *English Philosophy since 1900* (London: Oxford University Press, 1958). The books by Urmson and Warnock contain detailed appraisals of Russell's logical atomism. Russell's logical atomism as well as his neutral monism and his theories about truth and induction are sympathetically discussed by D. J. O'Connor in Ch. 26 of his *Critical History of Western Philosophy* (New York: Free Press of Glencoe, 1964). *Rivista critica di storia della filosofia* 8 (2) (1953): 101–335, and several articles in *Philosophy* 35 (January 1960): 1–50, are devoted to Russell's philosophy, including Anthony Quinton's useful sketch of the development of Russell's ideas in epistemology and metaphysics, "Russell's Philosophical Development," 1–13.

#### LOGIC AND MATHEMATICS

- Of Russell's own works on logic and mathematics, see *Principles of Mathematics* (Cambridge, U.K.: Cambridge University Press, 1903; 2nd ed., London, 1937); *Principia Mathematica*, 3 vols., written with A. N. Whitehead (Cambridge, U.K.: Cambridge University Press, 1910–1913; 2nd ed., 1927); *Introduction to Mathematical Philosophy* (London: Allen and Unwin, 1919); and the papers "On Denoting" (1905), "Mathematical Logic as Based on the Theory of Types" (1908), "The Philosophy of Logical Atomism" (1918), and "Logical Atomism" (1924), all of which are reprinted in *Logic and Knowledge* (see above).
- On Frege's parallel work, see his *Grundlagen der Arithmetik* (Breslau, 1884), translated by J. L. Austin as *The Foundations of Arithmetic* (Oxford: Blackwell, 1950); and P. T. Geach and Max Black, eds., *Translations from the Philosophical Writings of Gottlob Frege* (New York: Philosophical Library, 1952).
- Important critical discussions of Russell's work occur in W. E. Johnson, *Logic*, Pt. II (Cambridge, U.K., 1922), Chs. 3 and 6; F. P. Ramsey, *The Foundations of Mathematics* (London, 1931), papers I and II; W. V. Quine, *From a Logical Point of View* (Cambridge, MA: Harvard University Press, 1953), essays I, V, and VI; and G. E. Moore, *The Commonplace Book of G. E. Moore, 1919–1953*, edited by Casimir Lewy (New York: Humanities Press, 1963), Notebook II, item 4, and Notebook V, item 13.
- On formal implication, see A. N. Prior, "The Theory of Implication," in *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* 9 (1963): 1–6. On simplifications of type theory, see Alonzo Church, "A Formulation of the Simple Theory of Types," in *Journal of Symbolic Logic* 5 (1940): 56–68, and Ludwik Borkowski, "Reduction of Arithmetic to Logic Based on the Theory of Types," in *Studio Logica* 8 (1958): 283–295.

#### ETHICS AND RELIGION

Russell's early views on ethics are in "The Elements of Ethics," Ch. 1 of *Philosophical Essays* (New York: Longman, 1910); it has been reprinted in *Readings in Ethical Theory*, edited by Wilfrid Sellars and John Hospers (New York: Appleton-Century-Crofts, 1952), pp. 1–34. The fullest statements of his later position are in Ch. 9 of *Religion and Science* (New York: Holt, 1935) and in *Human Society in Ethics and Politics* (New York: Simon and Schuster, 1955). There are critical

discussions of Russell's views in Lillian W. Aiken, *Bertrand Russell's Philosophy of Morals* (New York: Humanities Press, 1963); in Justus Buchler, "Russell and the Principles of Ethics," in *The Philosophy of Bertrand Russell* (see above); and in D. H. Monro, "Russell's Moral Theories," in *Philosophy* 35 (1960): 30–50.

Russell's earlier views on religion are in "The Essence of Religion," in *Hibbert Journal* 11 (1912): 46–62. His first published discussion of the arguments for the existence of God is contained in Ch. 15 of *A Critical Exposition of the Philosophy of Leibniz* (Cambridge, U.K.: Cambridge University Press, 1900; 2nd ed., London and New York, 1937). His later views are expounded in several of the essays in *Why I Am Not a Christian* (see above) and in Pt. II, Ch. 7, of *Human Society in Ethics and Politics* (see above). The BBC debate with Father F. C. Copleston (1948), "The Existence of God," is available in the British edition, but not in the American edition, of *Why I Am Not a Christian*, but it has been reprinted in *A Modern Introduction to Philosophy*, edited by Paul Edwards and Arthur Pap, 2nd ed. (New York, 1965), and in *The Existence of God*, edited by John Hick (New York: Macmillan, 1964). Several chapters in *The Scientific Outlook* (London and New York, 1931) and in *Religion and Science* (see above) contain criticisms of the attempts of certain scientists to derive theological conclusions from physics and biology. Russell's objections to the fideistic position are found in Ch. 12, Bk. 3, of *A History of Western Philosophy* (see above). His objections to William James's defense of religion are contained in Ch. 29, Bk. 3, of the same work and in Ch. 5 of *Philosophical Essays* (see above). Russell's views on religion are criticized in H. G. Wood, *Why Mr. Bertrand Russell Is Not a Christian* (London, 1928); C. H. D. Clark, *Christianity and Bertrand Russell* (London: Lutterworth Press, 1958); G. S. Montgomery, *Why Bertrand Russell Is Not a Christian* (New York, 1959); and E. S. Brightman's contribution to the Schilpp volume, "Russell's Philosophy of Religion," pp. 537–556.

#### SOCIAL AND POLITICAL THEORY

In addition to the works mentioned in the first section of the present entry, the following among Russell's books dealing with social and political questions have been influential: *Principles of Social Reconstruction* (London: Allen and Unwin, 1916); *Roads to Freedom: Socialism, Anarchism and Syndicalism* (London: Allen and Unwin, 1918); *The Problems of China* (New York: Century, 1922); *Power: A New Social Analysis* (London: Allen and Unwin, 1938); *Authority and the Individual* (London: Allen and Unwin, 1949); and *New Hopes for a Changing World* (London: Allen and Unwin, 1951). Ch. 17 of *New Hopes* contains a moving discussion of the problems of growing old and facing death. Russell's fullest discussion of Marxism can be found in *Freedom and Organization 1814–1914* (London: Allen and Unwin, 1934); as *Freedom versus Organization*, New York, 1934), which is in effect a history of the main social and intellectual forces of the nineteenth century.

#### OTHER WRITINGS

Philosophical discussions sooner or later crop up in most of Russell's writings. Some of his most delightful occasional pieces have been collected in *Sceptical Essays* (London: Allen and Unwin, 1927); in *In Praise of Idleness* (London: Allen and Unwin, 1935); and in *Unpopular Essays* (London: Allen

and Unwin, 1950). The last of these contains his "Auto-biography," which was first published in 1936. *Bertrand Russell Speaks His Mind* (London: Barker, 1960) is a most interesting volume containing the unedited text of a series of television interviews, dealing with a great variety of topics, which took place in the spring of 1959.

*The Basic Writings of Bertrand Russell, 1903–1959*, edited by R. E. Egner and L. E. Denon (New York, 1961), is a very useful anthology of writings by Russell. The Schilpp volume contains an extremely comprehensive bibliography up to 1944.

#### RECENT WORKS

- Coffa, Alberto. "The Humble Origins of Russell's Paradox." *Russell* 33 (1979): 31–37.
- Eames, Elizabeth R. *Bertrand Russell's Theory of Knowledge*. London: George Allen and Unwin, 1969.
- Griffin, Nicholas. *Russell's Idealist Apprenticeship*. Oxford: Oxford University Press, 1991.
- Hylton, Peter W. *Russell, Idealism, and the Emergence of Analytic Philosophy*. Oxford: Oxford University Press, 1990.
- Irvine, A. D. "Epistemic Logicism and Russell's Regressive Method." *Philosophical Studies* 55 (1989): 303–327.
- Irvine, A. D., ed. *Bertrand Russell: Critical Assessments*. 4 vols. London: Routledge, 1999.
- Jager, Ronald *The Development of Bertrand Russell's Philosophy*. London: George Allen and Unwin, 1972.
- Kaplan, David. "What Is Russell's Theory of Descriptions?" (1970). In Pears (1972), 227–244.
- Landini, Gregory. *Russell's Hidden Substitutional Theory*. New York and Oxford: Oxford University Press, 1998.
- Linsky, Bernard. *Russell's Metaphysical Logic*. Stanford: CSLI, 1999.
- Lycan, William. "Logical Atomism and Ontological Atoms." *Synthese* 46 (1981): 207–229.
- Pears, David F. *Bertrand Russell and the British Tradition in Philosophy*. London: Collins, 1967.
- Pears, David F., ed. *Bertrand Russell: A Collection of Critical Essays*. Garden City, NY: Anchor Books, 1972.
- Putnam, Hilary. "The Thesis that Mathematics Is Logic." In Schoenman (1967), 273–303.
- Ryan, Alan. *Bertrand Russell: A Political Life*. New York: Hill and Wang, 1988.
- Sainsbury, R. M. *Russell*. London: Routledge, 1979.
- Savage, C. Wade, and C. Anthony Anderson, eds. *Rereading Russell: Essays on Bertrand Russell's Metaphysics and Epistemology*. Minneapolis: University of Minnesota Press, 1989.
- Schoenman, Ralph, ed. *Bertrand Russell: Philosopher of the Century*. London: George Allen and Unwin, 1967.
- Schultz, Bart. "Bertrand Russell in Ethics and Politics." *Ethics* 102 (1992): 594–634.

**Paul Edwards (1967)**

(*Life and Social Theories, Ethics and Critique of Religion*)

**William P. Alston (1967)**

(*Epistemology and Metaphysics*)

**A. N. Prior (1967)**

(*Logic and Mathematics*)

*Bibliography updated by Ernest Sosa (2005)*

## RUSSIAN PHILOSOPHY

In the broad sense the words “Russian philosophy” refer to all schools of philosophical thought pursued in Russia, regardless of differences among them. In the narrower sense the terms describe the religious-philosophical trend that flourished in the late nineteenth and early twentieth centuries. Both uses have value: The first embraces the variety of interests among Russian philosophers, whereas the second points to their most distinctive contribution to philosophy in general. But even on the broadest level, the common preoccupations that were typical of major Russian thinkers shaped the physiognomy of Russian philosophy as a whole.

Philosophy in Russia developed in a variety of forms. Philosophical ideas permeated religious, political, and literary debates throughout the country’s history. For a long time they were not articulated in what counted as philosophical parlance in the West, largely because of unfavorable historical conditions. But when these conditions changed, as they did, for example, in the late nineteenth and especially in the early twentieth centuries, there emerged a vibrant philosophical scene. This flourishing had been prepared within Russian culture, among other things, by its religious, literary, and scientific thought. Thus it should not be surprising that some theologians, novelists, and scientists are relevant to the history of Russian philosophy.

### TYPICAL FEATURES

For various reasons Russian philosophy has been dominated, not to say oppressed, by pragmatic concerns. Realistic or utopian, philosophical thought in Russia is expected to be engaged. It is not an accident that Marxism, for which social practice is the criterion of theoretical truth, has had such a firm grip on the Russian polity. Even when Russian philosophy did reach the heights of speculation—as in the thought of Vladimir Solov’ev (1853–1900)—it still bore the mark of “theurgic restlessness,” in Vasilii Zen’kovskii’s (1881–1962) words—that is, the desire to transfigure life. Still, when conditions were right, and sometimes despite harshly adverse conditions, Russian thinkers have achieved reflexive insights of uncommon depth.

Closely related to this is Russian philosophy’s realist *ontologism*; that is, the tendency to value the reality of *being* over and above the truths of abstract *understanding*. Nikolai Berdyaev (1874–1948) noted that the Russian mind strongly doubts whether the creation of culture is justified in the face of life’s problems. This doubt was typ-

ical of Lev Tolstoy (1828–1910) who disparaged art in contrast with the peasant’s work. Paradoxically, this tendency was also responsible for the seriousness with which Russians have treated the arts and philosophy. Likewise Russian thinkers often sought *justice* more eagerly than *truth* because the former seemed more tangible and urgent than the latter.

Many commentators have insisted that Russian philosophy is also inherently religious and personalistic. While the aggressively atheist and collectivist Soviet Marxism is an inescapable counterexample, it cannot be denied that the themes of religion and personhood have occupied and continue to occupy a prominent place in Russian philosophical discourse. Fëdor Dostoevsky’s (1821–1881) persistent interest is only the more familiar, especially to the West, among many manifestations of these themes.

Russian thought has a marked predilection for viewing things holistically. Russian philosophers have often been preoccupied with global, wide-ranging problems and visions of all existence as an integral whole. In metaphysics this trait is responsible for Solov’ev’s doctrine of all-unity. On the opposite end, this holism transmogrified into totalitarianism for which Stalinism stands as the most ominous example.

The evolution of philosophical ideas in Russia has been shaped by the persistent Slavophile-Westernist dichotomy; that is, tension between the impulses, on the one hand, toward national uniqueness and, on the other, toward closer affiliation with the West. However, from the earliest time these tendencies were so closely intertwined with each other that any attempt at a simple delineation is misleading.

And, finally, there is in Russian thought what Berdyaev called the “eschatological” orientation that can also be described as striving toward limits—in particular, the limits of thinking and of intelligibility of things. Like all the other features, this one also has had two opposite consequences. On the one hand, it makes Russian thought philosophically inclined in general, for it pushes rational enquiry to dwell persistently on ultimate questions. On the other hand, such a passion for limits could encourage, as it did in Berdyaev’s own case, impatience with careful argumentation.

### HISTORICAL EVOLUTION

Russian philosophical thought cannot be properly understood apart from its historical development. Its constant and eager immersion in cultural, social, and political con-



texts, as well as its stubborn continuity, make a historical perspective necessary for grasping both the problems that it grappled with and the solutions that it proposed.

**KIEVAN PERIOD.** Philosophical ideas, properly so called, first appeared in Russia when Christianity was introduced in 988 by the Kievan Prince Vladimir. The prior, polytheistic view of the world was partially replaced with the Christian outlook, resulting in the fertile amalgam of Eastern Orthodoxy and Slavonic paganism called “dual faith” (*dvoeverie*).

Universities and academic philosophy did not appear in Russia until the eighteenth century, nor was there a direct engagement with ancient Greco-Roman thought of the sort that shaped western medieval learning. Nascent Russian literature absorbed from Byzantium a number of early patristic writings, particularly those of the Cappadocian Fathers, in the form of religious-dogmatic texts translated from Greek into Church Slavonic. Anthologies comprising the writings of John Chrysostom, Gregory of Nazianzus, Basil the Great, Gregory of Nyssa, and John of Damascus introduced Russians to Christian Neoplatonist cosmology, metaphysical anthropology, and allegorical exegesis. Kievan Rus also imported the veneration of Sophia Divine Wisdom that found expression in architecture, icon-painting, and hymns.

In the mid-eleventh century this learning began to bear fruit when the first Russian Metropolitan of Kiev Ilarion described in his “Sermon on Law and Grace” history in terms of contrast between the law of the Old and the grace of the New Testaments, and argued the equal standing of Kievan Rus among Christian nations. Moral ideas were disseminated through “instructions” (*poucheniia*) for righteous living that often contained philosophical ideas derived from ancient and Byzantine thought. Throughout the premodern period philosophy in Russia was viewed primarily as *ancilla fidei* and a path toward religious illumination. At the same time it was understood in broad terms: Plato, Fathers of the Church, and even certain icon painters were considered “philosophers.”

Around the mid-thirteenth century this early flourishing was interrupted by the Mongol invasion. Bishop Serapion, who witnessed the sack of Kiev in 1240, was a proponent of the view that history was a series of catastrophes visited by God upon humanity for its sins. With Kiev devastated by the invasion, the center of religious and cultural life shifted to Vladimir and Moscow in the forested northeast that was less vulnerable to attack from the steppes.

**MUSCOVITE PERIOD.** In the fourteenth century the influence of hesychasm became pronounced, especially through the activities of St. Sergii of Radonezh (1314/22–1391/92). The Trinity-Sergius Monastery near Moscow that he founded soon rivaled Kiev’s Monastery of the Caves as Russia’s main religious center. St. Sergii’s popularity and influence signaled the rebirth of Russian culture around the Grand Duchy of Moscow that in 1380 successfully challenged the Mongol rule. The icon painter Andrei Rublev (d. c. 1430), whose art had a marked contemplative quality, was another representative of this cautious revival. In 1371 the translation of the Areopagitic corpus appeared that had a lasting impact on medieval Russian thought. (More than seventy copies of this work dating to the sixteenth and seventeenth centuries are extant.)

Russia’s final emancipation from the Mongol yoke followed soon after Byzantium’s fall in the mid-fifteenth century to Ottoman Turks. Both events affected Russians deeply. Soon monk Filofei proposed that Moscow was the third Rome (after Rome proper and Constantinople) and “there will be no fourth” (cf. Zen’kovskii 1953, pp. 34–35). The idea resonated with Muscovite rulers who sought to establish themselves on the European scene. According to this doctrine, Byzantium had fallen because it departed from the true faith and Russia now inherited its mission.

Two major debates convey the atmosphere of the time. Led by the hesychast Nil Sorskii (1433–1508), the so-called “Nonpossessors” (*nestiazhateli*) condemned accumulation of wealth by monasteries and sumptuous church ritual. Their opponent Iosif Volotskii (1439–1515) argued for economically strong monasteries that could help the unfortunate and have a part in social and political affairs. Nil Sorskii was, incidentally, among the first in Russia to refer to the “natural rights” of a person—a theme that gained currency in sixteenth-century religious and political polemics there. (As peasants were being enserfed, some religious writers argued passionately—but to no avail—against slavery as a violation of Christian principles.) The other dispute was the correspondence between Ivan the Terrible and Prince Andrei Kurbskii. The latter argued in favor of a sustained role of traditional aristocracy in government. The tsar’s course, however, was to assert his absolute authority with the help of a new gentry that completely depended on his favor. Conducted with ostentatious cruelty the policy did solidify Ivan’s autocracy but at a price: By the end of the sixteenth century Russia was in the throes of a major crisis.

On Nil Sorskii's side was Maksim the Greek (1470–1556), the most remarkable intellectual in Russia during that period. Originally a humanist scholar in Florence who later took monastic vows, he was invited in 1518 from Mount Athos to Moscow to assist in translating theological works. While a controversial figure for Russian ecclesiastical authorities, Maksim was nonetheless a scholar of European stature who helped spread philosophical knowledge in Russia.

With the seventeenth century came the painful “time of troubles”: Russia's medieval complexion began to change into a modern one. By the middle of the century political and religious tensions erupted in a major schism (*raskol*), which resulted in the separation from the Church of a large group of the so-called “Old Believers.” Given the western leanings of their opponents, the energetic Patriarch Nikon and Tsar Alexis (reigned 1645–1676), the schism has been viewed as a struggle between medieval fideistic and modern rationalistic outlooks.

Early in the century Petr Mogila established a spiritual academy in Kiev, fashioned after Polish (Jesuit) models. Secular schools began to appear in Moscow and in 1678 the first institution of higher learning was founded there: the Hellene-Greek Academy. The curricula of these schools included logic, psychology, and physics. The budding academia was occupied by the controversy between the “Graecophiles” faithful to the Byzantine roots of their learning and the “Latinists” influenced by western scholasticism.

**THE ENLIGHTENMENT.** Inaugurated by the reforms of Peter the Great (reigned 1696–1725), the eighteenth century became the time of a rapid assimilation of western European thought. Philosophical ideas from Europe were absorbed along with progress in the arts, secular education, and science. With the establishment of the Academy of Sciences and universities philosophy attained an official secular status. From translation, publication, and dissemination of foreign literature in the beginning, Russian Enlighteners eventually moved to creating their own works.

The most urgent task for the new educated elite was the development of a secular national ideology. The medieval ideal of “Moscow the Third Rome” was being replaced with the secular ideal of the Russian Empire. The first modern Russian historian, Vasilii Tatishchev (1686–1750) saw, in the Hobbesian vein, the basis of monarchy as the agreement between the sovereign and his subjects rather than in the sovereign's divine right. He

argued, in the proto-utilitarian spirit, that “the desire of well-being is inexorable in man and stems from God.” (Zen'kovskii, p. 79). His younger fellow-historian, Prince Ivan Shcherbatov (1733–1790), sharply criticized the established church—even as his political sympathies remained on the side of landed aristocracy. Tatishchev and Shcherbatov differed on the most burning moral question of that era, the freedom of the serfs, but both saw the well-being of the nation, rather than its religious mission, as the chief goal of the state.

The ideas of the Encyclopaedists circulated widely among the educated Russian society. Empress Catherine the Great (reigned 1762–1796) was an attentive reader of Charles Montesquieu's treatise *L'esprit des lois* (The Spirit of the Laws, 1748) and maintained correspondence with Voltaire, Diderot, and d'Alambert. Her friendship with *philosophes* doubtless stimulated Voltaire's near-cult status among educated Russians. Unavoidably this interest had much to do with a facile imitation of the West but it also had its serious side. Playing the part of an enlightened monarch, Catherine undertook a relatively progressive, if halting and ultimately unfinished, governmental and legal reform.

The accelerated development of the arts and sciences in this period was epitomized by the polymath and poet Mikhail Lomonosov (1711–1765). A fisherman's son from a northern province, he became the first Russian scientist of European stature and was instrumental in promoting scientific research and higher education in his country. To him belonged the famous prophecy, in verse, that combined the zeal of an Enlightener with national pride: “The Russian land can give birth to its own Platos and quick-witted Newtons.”

Simultaneously the traditional line of Orthodox theology was carried on by Paisii Velichkovskii (1722–1794) and St. Tikhon Zadonskii (1724–1783). Velichkovskii was a spiritual elder, the type best known from Zosima, a character in Dostoevsky's novel *The Brothers Karamazov*. Arguing against the alleged sanctification of the created world in secular thought, St. Tikhon taught that the external world had to be transfigured rather than accepted on its own terms. Concentrated on righteous living and one's personal connection with the Absolute, this theology was a welcome reprieve, as Zen'kovskii notes, from the burden of justifying Russian state messianism.

A counterpoint woven of both secular rationalism and religious mysticism was created by the most remarkable philosopher of the Russian eighteenth century, the Ukrainian Grigorii Skovoroda (1722–1794). A “Nonpos-

essor” and itinerant philosopher of a Socratic mold, Skovoroda expounded an original doctrine that was inspired by ancient sources, patristic thought, and modern European philosophy. There was a Christian Neoplatonist note in his belief that man’s proper purpose was an “erotic” ascent to divinity, as well as in his self-written epitaph: “The world tried to catch me, but has failed.” His influence on the contemporary philosophical scene was, sadly, almost nonexistent; his works were not published during his lifetime and began to attract serious attention only in the nineteenth century.

A different quest for spirituality outside the Church was evident in the movement of Freemasons that started in Russia in the second third of the eighteenth century. In the 1770s there emerged among them a group led by Nikolai Novikov (1744–1818) and Johann Schwarz (1751–1784). Novikov’s contribution was mostly as an editor and publisher: from 1779 to 1792 he published almost nine hundred titles that included, aside from Russian authors, translated works of Jacob Boehme, Voltaire, John Locke, G. E. Lessing, and Novikov’s favorite, Blaise Pascal. These Freemasons combined respect for natural science with the primacy of morality over the intellect.

Alongside modern scientific realism the nascent Russian intelligentsia absorbed western utopianism. As in the West, however, utopia often served as a vehicle for social criticism. Vasilii Trediakovskii in his *Tilemakhida* (1766), a verse translation of François Fénelon’s novel *Les Aventures de Télémaque*, described the torment of monarchs in Tartarus: they looked at their own monstrous images in the “mirror of truth.” From Trediakovskii’s poem came the epigraph to Aleksandr Radishchev’s (1749–1802) *Journey from St. Petersburg to Moscow* (1790): “The monster is opulent, impudent, enormous, hundred-mouthed, and barking.” The main target of Radishchev’s moral sermon from the standpoint of natural rights was the inhumanity of the institution of serfdom. It was the most striking fruit of the Russian Enlightenment, and cost the author dearly: he was exiled to Siberia. Novikov was likewise imprisoned in St. Petersburg. Frightened by the French Revolution, the aging Empress was now perturbed by the liberties her subjects were taking.

The turn from the Enlightenment to conservatism among Russian intellectuals was vividly exemplified by the historian and writer Nikolai Karamzin (1766–1826). A proto-Westernist, he was originally attracted to Locke and Rousseau but his views evolved from a vague empiricism and tolerant sentimentalism to defending the expe-

dience, for the stability of the state, of “enslaving people rather than prematurely freeing them.” The French Revolution was the key factor in this striking change. Karamzin initially hailed it as “the triumph of reason” but then, as terror struck, condemned it as the collapse of the Enlightenment. He was among the first to give Russians a serious perspective on their own history. The poet Aleksandr Pushkin (1799–1837) compared his discovery of Russia’s past to Columbus’s discovery of America.

The Russian Enlightenment drew to a close when, after Catherine’s death, Novikov was freed only to live out the remainder of his life in obscurity, and Radishchev, a few years after his release, committed suicide. But its ideas became an integral part of Russia’s intellectual makeup. Its complex legacy contained mutually intertwined, conflicting themes, such as national identity and universal humanism, secularism and religious tradition, scientific cognition and mysticism, art and morality, theoretical quest for truth vis-à-vis social practice.

**THE GOLDEN AGE.** Although rooted in a long-standing cultural and spiritual tradition, Russian philosophy proper was born in the nineteenth century. As it matured, it underwent several waves of foreign influence: idealist (especially German) in the 1830s and 1840s, positivist in the 1860s, Marxist in the 1880s and 1890s—to mention only the most poignant ones. Once it appeared, each strand remained an active factor in the continuing philosophical debate. Russian mentality has been described as inclined toward extremes, and the reception of Western ideas in Russia bears out this observation: their assimilation often meant radicalization. This was true of the “Nihilists” of the 1860s who developed a cult of natural science, and later of Vladimir Lenin (1870–1924), who stripped Marxism down to its bare essentials and ruthlessly pursued his vision. Solov’ev, by contrast, strove to synthesize diverse strands into a holistic idealist vision.

The famous phenomenon of the intelligentsia arose in this century. Recruited mostly from the middle class, the new educated elite developed a degree of self-consciousness one rarely finds in its Western counterparts. The idea of its “debt to the people,” articulated in Petr Lavrov’s (1823–1900) *Istoricheskie pis’ma* (Historical letters, 1868–69), shaped the ethos of this group. From the very beginning, though, the intelligentsia was torn by internal conflict and contradictions. Its admirers saw in it the “conscience of the nation,” its critics an intolerant “monastic order” of political radicalism, and many of its members were convinced that the two were synonymous. In the meantime such major thinkers as Solov’ev, Dosto-

evsky, and Tolstoy resisted being included among its ranks. In the early twentieth century philosophers of religious orientation subjected the intelligentsia's atheist outlook to an unflattering critique. They were noisily rebuked both by radicals and liberals. Left-wing intelligentsia played a crucial part in bringing about the revolutionary turmoil of the early twentieth century—the turmoil that led to its own dispersal in the thin air of history. Originally the flag-bearer for social progress and against despotism, in the Soviet period it became an evanescent specter. Its relation to the so-called “Soviet intelligentsia” was too problematic to warrant a continuum between them.

Early developments in philosophical education were not auspicious. Organized on Wolffian principles, academic philosophy had enjoyed steady growth since the middle of the eighteenth century. From 1817 and until the mid-nineteenth century, however, it suffered from a crisis precipitated by a conservative turn in Alexander I's policy and then exacerbated by the oppressive rule of Nicolas I (reigned 1825–1855). The teaching of philosophy was abolished for long periods in gymnasias and universities. A senior official summed up the government's view of it: “Utility is doubtful, whereas harm is obvious” (Radlov, *Ocherk istorii russkoi filosofii* [Essay on the history of Russian philosophy], 1920, p. 7). To circumvent restrictions some professors taught philosophy under the guise of other disciplines, such as history or geology. Philosophical instruction continued uninterrupted, however, in religious seminaries and academies but it was not until the second half of the century that the situation of academic philosophy began to be more or less normalized. Yet even as conditions improved, Russian thought retained much of its nonacademic character. For various, mostly political, reasons prominent thinkers—be it Aleksandr Herzen (1812–1870), Solov'ev, or Nikolai Chernyshevskii (1828–1889)—worked outside universities.

In the 1820s the first philosophical circle appeared; its members called themselves by the Russian equivalent of *philosophes*—*liubomudry*, “lovers of wisdom.” The group's leader, Prince Vladimir Odoevskii (1804–1869), presented a Schellingian view of Russia's future in his utopian dialogue-novel *Russian Nights* (1844) in which he gave a modern version of Russian messianism. History moved, he rhapsodized, toward “a holy triunity of faith, science, and art.” Anticipating Dostoevsky, he claimed that Russia was destined to accomplish this universal synthesis because of her “all-embracing multifaceted spirit.”

Such optimism, however, was in sharp contrast to the somber skepticism of Petr Chaadaev's (1794–1856) *Philo-*

*sophical Letters*. Chaadaev saw the West as the ideal of civilization; all other societies were, in his opinion, mere approximations to it, with Russia falling outside the category altogether. Chaadaev's bitterness was cast against the background of two recent events: Russia's victory over Napoleon in 1812 that encouraged hopes for the nation's greatness, and the crushing defeat of the 1825 Decembrists's uprising that extinguished hopes for reform and liberty. He was inspired in large part by Joseph de Maistre and Friedrich Schelling. He later fine-tuned his position to argue that Russia was called upon to resolve the contradictions that still plagued the West. The evolution of Chaadaev's views became typical for Westernists: from adulation of the West to disillusionment to seeing Russia's potential in her backwardness. The conviction that Russia was a “virgin soil” whose lagging behind could be turned to advantage as “the possibility of choice” became the cornerstone of Westernist constructions from Herzen to Vladimir Lenin (1870–1924). Chaadaev's caustic but profound outburst brought into existence two opposite trends, the Westernists and Slavophiles, whose mutual rivalry has since shaped, and continues to shape, the evolution of Russian thought.

Because of their intertwined destinies “Slavophiles” and “Westernists” come close to being the worst misnomers in the history of Russian thought. Both groups were deeply dissatisfied with the current conditions in Russia. Contrary to the xenophobic connotation of their name, many Slavophiles respected European learning and culture and kept abreast of recent Western philosophical thought. For their critique of the West they often borrowed ammunition from the West itself. Conversely, the Westernists' professed cause was to save Russia, and many of them even believed, such as Herzen, that Russia held the key to saving the West from the West's own woes. For both, the goal of “enlightening” Russia was of paramount importance, although they were divided on the possibility of “national science.” Slavophiles defended the idea (without defining it clearly), whereas Westernists rejected it in favor of universal rationality.

And yet their differences were not trivial. Slavophiles believed that, enviably advanced as it was, Europe had come to an impasse and Russia had to avoid a similar fate. The West's original sin, according to Slavophilism, consisted in the rationalistic tendency of Roman Catholicism that was codified in the *filioque*; that is, the dogma that the Holy Spirit proceeded from *both* the Father *and* the Son. Both early Slavophiles, such as Aleksei Khomiakov (1804–1860) or Ivan Kireevskii (1806–1856), and their later followers, such as Sergei Bulgakov (1871–1944) or

Nikolai Losskii (1870–1965), accused Catholic theology of replacing the mystery of the Holy Trinity with a hierarchical scheme in which the Holy Ghost was subordinated to the other two persons. This eventually led, via scholasticism, to Protestantism and thence to modern secularism. The decline of the authority of the Church in turn weakened, Slavophiles believed, the foundations of communal life and created the West's atomistic individualism. Russia, they claimed, could offer an alternative because its culture still contained the original wholesome elements, unspoiled by the westernization of the previous two centuries.

Against rationalism in epistemology the Russian mind could offer, Kireevskii argued, the ideal of integral knowledge in which rational thinking and divine revelation would be properly balanced. Against individualism in social philosophy it could offer *sobornost'*—the concept that amalgamates “togetherness” with “conciliarism” (from “church council”) and projects the ideal of an humanity united by love and faith, where the freedom of the individual is in harmony with the common cause. Khomiakov found its manifestation in the Orthodox Church and Konstantin Aksakov (1817–1860) in the Russian village commune. Russia's historical task was understood as universal, although it remained unclear how other nations, who had their own traditions, were supposed to accept Eastern Orthodoxy. Slavophiles' concern, however, was to outline Russia's potential place in the “family of nations” rather than to develop a specific strategy for attaining it. The mankind of the future was perceived, in Aksakov's terms, as a “choral person”—the notion that in the twentieth century was assimilated by Lev Karsavin (1882–1952) into his doctrine of humanity as a “symphonic person.”

Westernists, on the contrary, insisted that Russia needed to join advanced European nations in pursuing economic, social, and political progress. Where Slavophiles envisioned *sobornost'*, Westernists insisted on the legal rights of the individual. If Slavophiles found pristine purity in pre-Petrine Russia, Westernists blamed the country's slow progress on xenophobic medieval Russian tsardom. Their sharpest difference from Slavophiles, however, consisted in their hostility toward religion. In Herzen's words, there was an “ecclesiastic wall” between him and his opponents. The common limitation of both was their utopianism: One idealized Russia's past and the other, the West's present. Furthermore, for neither of them philosophy had independent value but was merely an instrument for achieving goals other than knowledge and understanding.

The reception of Schelling and Hegel casts a helpful light on the manner in which philosophy's tasks were conceived. Schelling's philosophy enjoyed a warmer reception—at least in the religious segment of Russian thought. In fact, there is some truth to Arsenii Gulyga's (1921–1996) remark that “Russian philosophy is a Schellingian.” With Hegel Russians tended to distance themselves, even as they respectfully learned from him; in Schelling they found a kindred spirit. The view of the world as an organic whole has had more followers and fewer detractors in Russia than in the West; it retains importance there to this day. Schelling's doctrine of intellectual intuition proved particularly attractive to Russian thinkers. From Odoevskii to Solov'ev they embraced the notion of an immediate meeting of consciousness with both inner and outer reality; in the twentieth century it inspired a whole intuitivist school. Chaadaev was deeply affected by Schelling's philosophy of revelation; Kireevskii and Solov'ev, by his epistemology; Odoevskii and Bulgakov, by his *Naturphilosophie*; and Aleksei Losev (1893–1988), by his aesthetics and philosophy of myth. Many of them found in Schelling's thought inspiration for viewing art and religion as (extrarational) sources of rational thinking.

Russian liberal thought was, by contrast, at its inception primarily Hegelian. Vissarion Belinskii's (1811–1848) and especially Herzen's engagement with Hegel's philosophy were typical. Both embraced Hegelianism in the beginning but then rejected what they perceived as its abstract universalism. Belinskii, on the one hand, got most of his Hegel via Mikhail Bakunin (1820–1900) who at the time was an overenthusiastic Hegelian. Herzen, on the other hand, attentively studied Hegel's writings firsthand. The result was, however, more or less similar. “(Hegelian) reason does not know,” Herzen impugned, “*this* person but only the necessity of a person in general ...” (*Zen'kovskii*, pp. 285–6). The main point of Herzen's dissatisfaction was the same as Karl Marx's: life is not merely about thinking, he insisted, but chiefly about acting in the world. Virtually all Russian philosophers turned away from Hegel upon initial acquaintance. Those consumed by revolutionary causes, such as Bakunin, blamed him for excessive contemplativeness, whereas Slavophiles and religious philosophers rejected his doctrine of a rationally cognizable absolute. Various parts of Hegel's system were adopted but only the rarest exceptions, such as Boris Chicherin (1828–1904), accepted its essential core, the doctrine of the absolute concept. Characteristically, Herzen found in Hegel's dialectic “the algebra of revolution”—a description that was later eagerly endorsed by Lenin. This appropriation

epitomized the political pragmatism that was imposed on the German philosopher's speculative method.

Philosophers' concerns for "the concrete person" were nourished by the burgeoning Russian realist literature that paraded, in an intensely empathetic light, a series of characters whose suffering was a condemnation of a social order in which human dignity was out of place. Conversely, Russian thinkers frequently offered their insights in literary form. In fact, the most burning of the "cursed questions" that preoccupied the intelligentsia throughout its existence were articulated as titles of literary works: Herzen's 1847 novel *Who Is to Blame?* and Nikolai Chernyshevsky's (1828–1889) 1863 socialist utopia *What Is to Be Done?* The latter query proved particularly haunting: Leo Tolstoy in 1883 and Lenin in 1902 each wrote a work bearing similar titles.

Hegel and Schelling were soon replaced by Ludwig Feuerbach and Left Hegelians as socialist ideas were spreading among educated Russians. In the 1860s materialism propounded by Ludwig Büchner and others was added to the mix; it was embraced by the so-called "Nihilists" whose leading figures were Dmitrii Pisarev (1840–1868), Nikolai Dobroliubov (1836–1861), and Chernyshevskii. Pisarev's crude materialism, however, was not so much a philosophical position as a propagandistic means of destabilizing old religious and social values. Calculated to outrage, his maxim that "boots are more valuable than Shakespeare" was, in fact, a call to social activism as opposed to the aesthetic hedonism of the leisure classes. It was also a message about the utility of science and technology; that is, the business of the newly emerging class of physicians and engineers, contrasted with the aristocratic art of the previous era. The most articulate thinker of the "Nihilist" camp, Chernyshevskii, by contrast, argued for genuine art that would be a life-transforming praxis rather than idle entertainment. The rise of Nihilism marked the radicalization of Herzen's intellectually broad and humane liberalism, and the beginning of the latter's transfiguration into fanatical revolutionism.

In the late 1860s and 1870s the earlier materialism was absorbed into the broad social, cultural, and ideological movement called "Populism" (*narodnichestvo*). Its intellectual leaders, Lavrov and Nikolai Mikhailovskii (1842–1904), combined positivist epistemology and materialist metaphysics with an evolutionist view of history. The Populists' goal was socialism in Russia, on the basis of the village commune. Their views about both the goal and the ways of achieving it, however, varied from the anarchism of Bakunin and Petr Kropotkin

(1842–1921) to the conspiratorial terrorism (with a Marxist tinge) of Petr Tkachev (1844–1886). The Populists' main philosophical difficulty consisted in reconciling the individual's agency with positivist determinism. Like their materialist predecessors, however, these thinkers did not embrace a particular philosophy of nature or history for its intellectual merits but were interested primarily in using it for social change. It was Mikhailovskii who pointed out, memorably, the conflation of "truth" and "justice" in the Russian word *pravda* that has since come to signify one of the most pervasive features of the Russian philosophical mindset. It was also Mikhailovskii whose "subjective method" in sociology was intended to enhance the ability of "critically thinking individuals," as Lavrov called them, to influence the course of history. Populism later evolved into the political party of Socialist Revolutionaries, the Bolsheviks' most powerful left-wing rival, and its ideas continued to exercise their influence well beyond its final collapse in the 1920s.

Less influential was the moderate liberal thought of such thinkers as Konstantin Kavelin (1818–1885) and Boris Chicherin (1828–1904) who defended, from an Hegelian position, the ideals of the law-governed state in political theory and the universal "higher synthesis" of religion and philosophy in epistemology. As the earlier Westernism was radicalized, so too the original, rather moderate Slavophilism was producing its own increasingly radical offshoots. Konstantin Leont'ev (1831–1891) offered a scathing critique, on aesthetic grounds, of contemporary Western society. Unlike Friedrich Nietzsche with whom he is frequently compared, Leont'ev ended not with a call for a proud Overman, but with a return to an ascetic Orthodoxy. Nikolai Danilevskii's (1822–1885) theory of "cultural-historical types" advanced a cyclical model of history in which the tired Romano-Germanic civilization was about to yield its place to a younger Pan-Slav one. Danilevskii's ideas had an impact on the "back-to-the-soil" group of authors (*pochvenniki* from *pochva*, the Russian for "soil"), whom Dostoevsky lent his not insignificant authority.

Dostoevsky was, incidentally, one of the first Russian thinkers who had a marked influence on Western philosophy. His explorations of the religious, moral, and psychological dimensions of the human condition made a deep impression on both contemporaries such as Nietzsche and later figures such as Albert Camus. Inside Russia Dostoevsky's ideas reverberated in the religious-philosophical school of the early twentieth century.

The more liberal patrimony of Slavophilism, however, was cultivated by Russia's first truly great philosopher Solov'ev. Solov'ev's philosophy was an impressive attempt to fuse together positivism, idealism, and mysticism. His early critique of positivism evolved into the assimilation of Auguste Comte's ideas into his own view of history as divine will unfolding toward "free theocracy." Comte's *Grand Être* was likewise absorbed, along with Gnostic, Cabalistic, Eastern Orthodox, and German Romantic ideas, into Solov'ev's neoplatonist metaphysics of Sophia Divine Wisdom. Later Solov'ev performed a similar operation on Chernyshevskii's positivist aesthetics by interpreting it in the light of his own doctrine of art as theurgy: that is, humanity's continuation of divine creation. Yet his syntheses were not eclectic but rested on a broad conceptual foundation and formed a more or less coherent system—the first created by a Russian philosopher. With his more eager, ecumenical acceptance of the West Solov'ev modified earlier Slavophilism and worked to reconcile it with Westernism. Above all, however, his most lasting contribution consisted in the apologia of philosophical idealism. Solov'ev and Dostoevsky remained lonely voices among the intelligentsia during their lifetime but by the time of Solov'ev's death a reaction had already begun among a new generation of philosophers against secular ideologies and in favor of a serious engagement with religion.

While the rebirth of philosophical idealism was only dawning, however, its antipode was vigorously gaining ground. Marxism was known in Russia since the late 1840s but in its early stages it was only one among several currents of socialist thought. Nevertheless, it soon attracted significant interest: In 1869 Bakunin published (abroad) his translation of the *Communist Manifesto*, and three years later Russian became the first foreign language in which the first volume of *Das Kapital* appeared. By the end of the century Marxism became the most influential political doctrine among the intelligentsia. It established itself in competition with earlier socialist theories, primarily Populism. In contrast to Populists who wished Russia to avoid capitalism and leap, via village commune, directly into socialism, Marxists viewed capitalism as a stepping stone to socialist revolution. The abolition of serfdom in 1861 by Alexander II gave a strong impetus for the development of capitalist enterprise and, as the number of factory workers grew, socialist theorists began to pin their hopes on the new class. The key figure in the transition from Populism to Marxism was Georgii Plekhanov (1856–1918). His main concern seems to have been to elaborate a philosophical system based on Marxist precepts, while guarding the original doctrine against

misinterpretation and revisions. A significant feature of Plekhanov's reception of Marx's ideas was their refraction through Frederic Engels's work. Russian Marxists did not always take care to distinguish Marx from Engels and often argued—in fact, often they simply assumed—the unity of the two founders' respective positions.

In the last quarter of the century Russian academic philosophy finally became the key factor on the philosophical scene. The generation of Solov'ev and Mikhailovskii was receding into the past and most leading thinkers now taught at universities. Chicherin gradually developed his own system with an emphasis on the philosophy of right and of history. A Leibnizian revival was evident in the trend started by Aleksei Kozlov (1831–1931) that stimulated the development of personalism in Russian thought. The latter had an exceptionally far-reaching impact on such thinkers as Berdyaev, Losskii, and Lev Shestov (1866–1938). This was also the time when Kant's presence in Russian thought finally came to match that of Schelling and Hegel. The leading neo-Kantian Aleksandr Vvedenskii (1856–1925) concentrated on logic and philosophical psychology. Advocated by a number of scientists and philosophers, such as Vladimir Vernadskii (1863–1945) and especially Vladimir Lesevich (1837–1905), neopositivist thought was another major current in academic philosophy. It was concerned almost exclusively with the philosophy of science and empirical epistemology. Vernadskii's ideas later played an important part in what became known as Russian cosmism. The original tenets of this loosely defined trend were formulated by the (nonacademic) Nikolai Fedorov (1828–1903) whose eccentric hybrid of positivism and Christian eschatology aimed at the physical resurrection of all past generations.

**THE SILVER AGE.** The flourishing of the arts and philosophy, roughly, from 1890 to 1925 is often referred to as the "Silver Age." It was marked by the rise of Symbolist poetry, modernist music, avant-garde art, and a general invigoration of cultural life. The Silver Age unfolded against the background of growing capitalism and a relative liberalization of political life, punctuated by wars and revolutionary turmoil. New developments in the arts underscored expectations of tectonic shifts in political history. The theme of an impending catastrophe—hailed as a purifying storm by some and feared as a fatal calamity by others—haunted artists and philosophers alike. Russia's humiliating defeat in a war with Japan precipitated the first, abortive popular uprising in 1905. The tsarist government agreed to halfhearted parliamentary reforms but they were undermined by the outbreak of

World War I in 1914 and then annulled altogether by the Bolshevik revolution of October 1917.

Russian philosophy matured during this period. From the 1890s on government restrictions were loosening and in the early 1900s the autonomy of universities finally began to materialize. In 1889 the first professional philosophical journal, *Voprosy filosofii i psikhologii* (Questions of philosophy and psychology) was founded, followed in the first decade of the new century by several other publications specializing in philosophy. In 1897 the St. Petersburg University Philosophical Society was established and a few years later it was joined by the Religious-Philosophical Society in Memory of Vladimir Solov'ëv in Moscow and the Religious-Philosophical Society in St. Petersburg.

Contacts with European philosophy reached a high point. Russian philosophy was now fully integrated, if still as a minor partner, into the European philosophical culture. The most recent developments in Western thought were quickly assimilated by Russian thinkers; empirio-criticism and phenomenology were only the more notable among such new trends. The growing influence of Kant was mentioned above. Nietzsche's impact on the Russian thought of this period was profound and pervasive.

Two opposite, unequal trends dominated the scene during this time: Marxism and religious philosophy. The former was philosophically unimpressive but politically influential, whereas the latter, on the contrary, was politically insignificant but philosophically fertile. Their complex mutual interactions, ranging from antagonism to fusion, were the manifestations of a dynamic and visionary rather than rigorous *Zeitgeist*. Scientific positivism and political liberalism also continued, adding to the increasingly vibrant philosophical life.

The brand of Marxism that emerged as a result of Plekhanov's efforts and was now endorsed in the main by Lenin included the following basic components. It was founded on a materialist ontology; that is, the view that matter constitutes the source of all existence. Materialism was enhanced by a positivist epistemology that held modern science to be the only legitimate source of knowledge. Marxism considered itself a true—in fact, the only true—doctrine *because* it was a modern scientific theory. Its next key component, historical materialism, was the result of synthesizing the first two with Hegel's philosophy of history. And finally the whole was held together by dialectical materialism, also a permutation of Hegelian dialectics adapted to fit materialism and positivism. (Needless to say both Hegel's philosophy of history and

his dialectics were drastically deformed in these hybrids.) Materialist orientation also dictated that all social and political phenomena be viewed as determined by a society's economic base. The latter developed, according to the theory, over periods of gradually accumulating quantitative changes leading up to abrupt moments of revolutionary qualitative change. The result was the view that history was logical progress from one socioeconomic formation to another, culminating in communism as the most rational system. There was no room for divine authority in this picture; militant atheism was an indelible feature of Russian Marxism. In an apparent contradiction to its own economic determinism, the key factor in the "inevitable" socialist revolution was Marxism itself as a doctrine of "scientific socialism." Further, despite being the most revolutionary class, the proletariat had to be educated; as Lenin argued, "scientific socialism" had to be instilled in its consciousness.

In ethics universal moral values were rejected as products of "abstract bourgeois humanism" in favor of the view that all values were determined by class interest. The corollary was that, as the revolutionary vanguard of society, the proletariat held values that were superior to those of any other class. In aesthetics a similarly class-based criterion was adopted: judgment about art was determined by which class interest it promoted. Leo Tolstoy's oeuvre, for example, was famously described by Lenin as "the mirror of the Russian revolution." These principles received a less stark complexion once they were combined with a dialectical view of history according to which new eras partially reject but also partially absorb the achievements of previous ones. Thus the proletariat was supposed to have inherited the best that world civilization had developed prior to socialist revolution. But the ultimate authority on all issues belonged to the proletariat's own vanguard, the Communist Party. Likewise Lenin's unabashedly utilitarian, ideological aesthetic eventually replaced Plekhanov's earlier, more nuanced attitude as the official "partisan principle" in evaluating art.

There soon evolved two currents in Russian Marxism: radical and moderate. Lenin and his fellow Bolsheviks (the term "Bolshevik" literally means "a member of the majority") promoted the former, whereas the so-called "legal Marxists" that included Petr Struve (1870–1944), Berdyaev, and Bulgakov, advocated the latter. The radical trend absorbed from extremists such as Tkachev revolutionary voluntarism and justification of terror as a means of political change.



The main controversies that divided these currents had to do with whether Russia could bypass an extensive phase of capitalism and bourgeois democracy, and proceed directly to a socialist revolution. Lenin answered in a resounding affirmative, whereas his opponents, including Plekhanov, favored a less precipitous path. They feared that the dictatorship of the proletariat, which the Bolsheviks envisioned as the key instrument of transition from a semifeudal to socialist society, would be as oppressive as the tsarist regime. The question of eschewing a prolonged capitalist phase was also bound up with whether Russia could pursue the socialist path alone among nations. In classical Marxism progress toward socialism had been envisioned as an international process because capitalism was itself an international system, too entrenched for the proletariat of one country to overpower it. Russian Marxists split on the issue: The moderate wing laid stress on international cooperation and advocated waiting for ripe conditions in advanced European nations, whereas the radical wing insisted that it was possible to establish socialism in one country.

The main philosophical difficulty for Marxists stemmed from the materialist foundation of their doctrine and consisted in explaining how purely physical, unconscious matter could generate movement and, ultimately, consciousness. The argument that matter evolved in accordance with the laws of nature only raised questions about the origin of these laws themselves. Plekhanov and Lenin asserted that science disclosed what matter was but this claim lost its persuasiveness as new conceptions of matter were developed in physics and the hypothetical nature of these views became increasingly apparent. Lenin's statement that "matter is objective reality given us in sensations" was vague enough to accommodate idealism and thus created more problems than it solved. Similar problems haunted Marxist ethics. The critics of the dogmatic trend, such as Struve, complained that class interest did not provide a firm foundation for morality and, further, dissolved individual agency in socioeconomic forces. The dismissal of art as an activity with a distinct purpose was also problematic. Nor did philosophy itself fare better. "From Marx's and Engels' point of view," wrote Lenin in his essay "The Economic Meaning of Populism" (1894), "philosophy has no right to independent existence and its subject-matter divides itself [literally 'disintegrates,' *raspadaetsia*] among several branches of positive science."

In evaluating its claims, however, it is critical to realize that Russian Marxism was first and foremost a doctrine of political action. Its logic, philosophy of history,

social philosophy, epistemology, and even materialist ontology were adopted under the pressure of a specific sociopolitical ideal. It was the ideal of a strictly secular, modern society aimed at assuring the fullest realization of the immanent human potential by rationalizing the production and distribution of material wealth. The "superstructure" was to align itself with, and serve the achievement of, this goal. Hence Lenin's relentless defense of materialism, insistence on the scientific nature of Marxism, and uncompromising atheism. In Lenin's thought Russian Marxism's ideological pragmatism reached its apogee. Scant and unimpressive at best, his philosophical writings were all occasioned by topical debates and aimed at ensuring the resolve of the Bolshevik party. The motivation for his most extensive philosophical work, *Materializm i empiriokrititsizm* (Materialism and empiriocriticism, 1909), for example, was to rein in his comrades Lunacharskii and Aleksandr Bogdanov (1873–1928) who had strayed into "God-building" and "empiriomonism." The only exception was Lenin's *Filosofskie tetradi* (Philosophical notebooks, 1914) in which a more serious engagement with Hegel was evident, but these were private ruminations published only posthumously. Materialism and dialectics were meaningless for Lenin unless they were employed for the communist cause. "Materialism," he wrote, "includes partisanship (*partiinnost'*)." (*Collected Works*, Vol. 1, 1960, p. 401).

The moderate branch of Russian Marxism was more in earnest about resolving the philosophical difficulties of the doctrine but attempted solutions led to revisions of its original materialist, positivist, and deterministic tenets. The "legal Marxists" Struve, Berdyaev, and Bulgakov eventually abandoned orthodox Marxism in favor of philosophical idealism.

This was a sign of the opposite trend that became evident in the emergence of neo-Kantianism and especially religious idealism. The return to Kant was chiefly a development in academic philosophy, whereas the turn to religion swept along academics, independent thinkers, and artists. In later literature the appearance of a group of philosophers who drew inspiration from religion was described as a "religious-philosophical renaissance." The writer Dmitrii Merezhkovskii's (1865–1941) quest for a "new religious consciousness" was a more popular manifestation of this trend. Merezhkovskii initiated Religious-Philosophical Meetings in 1901–1903 as an attempt at a rapprochement between the church and the intelligentsia. The participating sides were ill at ease with each other and after Vasilii Rozanov's (1856–1919) character-

istically shocking call upon the clergy to sanctify physical sex the meetings were stopped on government's orders.

Three publications mark the evolution of this trend in the first two decades of the century. The 1902 anthology *Problemy idealizma* (Problems of idealism) was an initial attempt to revive idealism as a viable contemporary philosophical position, followed by the *Vekhi* (Landmarks, 1909), a cutting critique of the intelligentsia's ideological dogmatism, atheism, and social isolation, and, finally, by *Iz glubiny* (De profundis, 1918), a reaction to the Bolshevik revolution as an anti-Christian act prepared by the spiritual, cultural, and moral crisis of the previous two decades. (This indictment was echoed by Rozanov who called the Revolution "the apocalypse of our time.")

The crucial problem that these thinkers confronted was the reconciliation of philosophy with religion. The impulse to embrace religion came as a result of recoiling from materialism and positivism. Many religious philosophers began as Marxists in their younger years and then underwent an idealist conversion. But an attempt to reconcile religion and philosophy led to the choice between fideism and rationalism. Like their predecessors, Slavophiles and Solov'ev, Russian religious philosophers ultimately leaned toward the former. Ernest Radlov (1854–1928) even claimed in his 1920 *Ocherk istorii russkoi filosofii* (A survey of the history of Russian philosophy) that the tendency among Russian thinkers toward a mystical solution of ethical and epistemological questions was a "national trait." A closely related task that these philosophers pursued was defense of idealism. In many cases such defense involved rethinking the relation between ideas and empirical reality and resulted in a number of constructs: the "concrete idealism" of Sergei Trubetskoi (1862–1905), "ideal-realism" of Losskii, and "mystical realism" of Berdyaev. The "abstract" thought of German Idealism often served as a contrasting foil for these attempts to bring idealism closer to life.

At the same time the religious-philosophical school argued for a secular culture and philosophy informed by the Orthodox faith—domains that had been neglected, in their opinion, by the Russian Orthodox Church. Florenskii's 1914 classic *Stolp i utverzhdenie istiny* (The pillar and ground of truth) was perhaps the most monumental attempt to fuse together a modernist philosophical and aesthetic sensibility with Orthodox faith. The most important sources of inspiration for them included the thought of the early Slavophiles and especially of Vladimir Solov'ev. In epistemology they questioned both extreme rationalism and extreme fideism but their atti-

tudes varied widely. In method their approaches ranged from Losskii's strict adherence to formal logic to Semen Frank's (1877–1950) moderate dialectics to Berdyaev's aphoristic impressionism. In metaphysics many of them followed and further developed Solov'ev's doctrines of all-unity and Sophia Divine Wisdom. Their views on philosophy of history encompassed Florenskii's admiration of the Middle Ages, at one pole, and Berdyaev's progressivist Christian socialism, at the other. In political philosophy they were likewise diverse: Ivan Il'in (1883–1954) rigidly advocated monarchism, whereas Viacheslav Ivanov (1866–1949) vaguely evoked mystical anarchism. The only thing that united them was the conviction that modern secularism had exhausted itself and the reinvigoration of philosophy and culture in general was to be sought in a union with religion.

A particularly notable contribution by this group was their writings on the history of Russian philosophy. Evgenii Trubetskoi's (1863–1920) classic study on Solov'ev, Berdyaev's essay on Khomiakov, Gustav Shpet's (1879–1937) hypercritical survey, and Radlov's work mentioned above were part of this self-examination. A special place in this literature belongs to works on the "Russian Idea." Rooted in the writings of Dostoevsky and Solov'ev, this trope grew into a body of literature created by several generations of philosophers. On the broadest level, it referred to the unique Russian type of consciousness, culture, historical destiny, and place among the peoples of the world. After the Revolution this tradition was further elaborated in Eurasianism and culminated in Berdyaev's classic *Russkaia ideia* (The Russian idea, 1946). It eventually reemerged in post-Soviet thought where it took on still other interpretive hues.

Along with metaphysical, epistemological, and political issues, an exceptionally preeminent concern for this group was art, which they viewed as a conduit for religious enlightenment. Evgenii Trubetskoi, Florenskii, Bulgakov, and Berdyaev all dedicated to art some of the most inspired pages of their philosophical prose. Their insights into icon-painting (which Trubetskoi described as "theology in color"), liturgy (which Florenskii interpreted as the Orthodox *Gesamtkunstwerk*), and artistic creativity in general remain to this day exemplary in their subtlety and depth.

Even more than before Russian philosophy evolved during this time in an intense dialogue with the arts. The Russian avant-garde was often inspired by, and inspired in turn, the volatile mix of philosophical ideas. Among artistic movements Symbolism stood out, both in terms of its artistic influence and engagement with philosophy.

Poets Andrei Belyi (1880–1934), Aleksandr Blok (1880–1921), and especially Ivanov keenly explored the philosophical dimensions of their art. Symbolists came to the view, rooted in Romanticism and Solov'ev's theurgy, that art provided access to the "more real" plane of being and was a path toward spiritual or even cosmic transfiguration. Both philosophers and artists were fascinated with the limits of art. A wide array of artistic movements was driven by a desire to break down the barrier between art and life raised by Kantian disinterested aesthetic contemplation. The pivotal event of this period, the Bolshevik Revolution, did not initially stop this feverish activity but marked a watershed that inaugurated a new phase in the history of Russian philosophy.

**THE SOVIET PERIOD.** Two major processes were under way in the 1920s: the decline of the Silver Age and the rise of Soviet ideology. The new government sought a total submission of philosophy to state ideology and the means by which this was assured ranged from administrative pressure to exile to physical annihilation of dissenting thinkers. Berdyaev's Free Academy of Spiritual Culture and the Free Philosophical Association founded by Radlov, Losskii, and others in St. Petersburg were short-lived attempts to continue prerevolutionary activity. Philosophers associated with both were expelled from the country in 1922 among a large number of thinkers and scholars unsympathetic to the Bolshevik regime. In 1921 the teaching of non-Marxist philosophy was banned and in 1923 philosophy was replaced by dialectical materialism in higher education.

The tasks of Soviet philosophy consisted in "developing" Lenin's patrimony (which meant strictly adhering to its key tenets), combating domestic and foreign "bourgeois idealism," justifying the Party's political decisions, and supplying methodology to the sciences. Formulated even before Soviet philosophy as such was in existence, these tasks remained unchanged throughout the Soviet period. As state ideology Soviet Marxism was based on the Plekhanov-Lenin interpretation of Marx and Engels's views that was soon branded "Marxism-Leninism."

The debate during the 1920s between the so-called "mechanists," such as Nikolai Bukharin (1888–1937), and "dialecticians," led by Abram Deborin (1881–1964), was "resolved" by a ukase from the Communist Party. The episode served to solidify the typical Soviet way of "philosophizing": The last appeal was not to logic and reason, but to the recorded opinion of the "classics of Marxism-Leninism." The highest authority in interpreting the latter belonged, in turn, to the leadership of the Party. The

debate highlighted the paradox encapsulated in the expression "Soviet philosophy." On the one hand, Soviet ideology was based on a *philosophical* theory; on the other hand, this theory was *dogmatically* accepted as the final word in all ultimate matters. As a result, Soviet philosophy was implicitly burdened with the impossible task of reconciling the internal contradictions of Marxism—but only by appeal to Marxist principles themselves. The basic contradiction of the doctrine consisted in the fact that it insisted on the ontological primacy of matter over spirit but at the same time wished to be a theory (i.e., spirit) that changed the material world.

The untenable nature of this exercise did not escape contemporaries. Losev, whose eight volumes published between 1927 and 1930 were the swan song of the philosophical Silver Age, publicly called dialectical materialism a crying absurdity and challenged Soviet Marxists to acknowledge that their professed scientific rationalism was at bottom as mythological as any theology. His was a lonely voice, however, and it was silenced forthwith by an arrest, confinement at labor camps, and a ban on publishing upon release.

The repressions of the 1930s were the lowest point in the history of philosophy in Russia. The pre-Soviet intelligentsia was either intimidated or physically annihilated. In 1937 Florenskii and Shpet were executed in the Gulag. Russia was being purged of its philosophy. To train the new cadre was the task of the recently established Institute of Philosophy in Moscow. There were some attempts to simulate philosophical activity but they were crude and tendentious beyond redemption. Stalin's chapter on dialectical materialism in the 1938 *Kratkii kurs istorii KPSS* (History of the Communist Party of the Soviet Union [Bolsheviks]; short course) was not a philosophical work; it merely sealed the reduction of philosophy to ideological indoctrination for which epistemological or logical concerns were irrelevant. During Stalin's time the voluntaristic (i.e., ultimately terrorist) component in Russian Marxism overshadowed its other aspects. The three-volume *Istoriia filosofii* (History of philosophy) that appeared in 1940 brought to a simplistic pitch a tradition of interpretation established already by Lenin. The entire history of philosophy was presented as a struggle between "progressive" materialism and "reactionary" idealism. In 1947 Georgii Aleksandrov (1908–1961) published his *Istoriia zapadnoevropeiskoi filosofii* (History of Western-European Philosophy), based on similar principles of analysis.

This *History* figured prominently in Andrei Zhdanov's speech the same year, in which he announced

to the new generation of philosophers the Party's orders to be "more creative." Zhdanov's admonitions had a certain positive effect: For the first time since the 1920s the history of Russian philosophy, for example, became a legitimate subject. To bolster Marxism-Leninism's pedigree Soviet authors ingeniously discovered materialism in the ideas of Russian thinkers. Radishchev, Herzen, and Belinskii were recruited into the ranks of Lenin's precursors. Pisarev, Dobroliubov, and Chernyshevskii were, somewhat more justifiably, painted as "revolutionary democrats" and their materialism as a spontaneous discovery of truth prefiguring "scientific socialism." Tendentious as it was, this work was a step forward from the previous period of forced oblivion.

In the meantime philosophers of non-Soviet orientation continued to write privately "into the drawer." After his release from the camps Losev wrote treatises on ancient mythology and aesthetics, as well as philosophical prose. His fellow-survivor from the Silver Age Mikhail Bakhtin (1895–1975) worked on his theories of literature and culture; and Vernadskii developed his doctrine of the noosphere. The ideas of these authors became known only decades later when their works contributed to the intellectual ferment of the 1960s–1980s.

After Stalin's death in 1953 and especially after Nikita Khrushchev's 1956 official condemnation of Stalin's "personality cult" a "thaw" began during which ideological constraints on philosophy were gradually loosened. Khrushchev made an attempt to boost slipping enthusiasm for communism by adopting a new program for the Party but the effect of its exorbitant promises was cynicism rather than renewed optimism. Leonid Brezhnev and the new generation of leaders who came to replace Khrushchev were even less capable of reviving the decaying ideology and from the late 1960s a period of ever deepening disillusionment set in that eventually led to the fall of the Soviet Union in 1991.

In the 1950s and 1960s Soviet philosophy became an increasingly complex agglomeration of disciplines and approaches. The list of permissible themes gradually expanded. The precept, for example, of sacrificing the individual to the needs of the socialist state began to be revised as the human person was cautiously explored as a philosophical subject. Debates on the nature of philosophy ended, thankfully, in an ambiguity as to whether it was a science, theory of action, or world view. The discussion of materialism and dialectics likewise led to a number of diverging positions that included even disagreement with Lenin. Restrictions were still in place and it was impossible to challenge official orthodoxy directly

but attempts to solve its problems objectively tended to water down and sometimes even to dissolve its basic precepts. Some philosophers sought refuge from ideology in such relatively neutral areas as philosophy of science, logic, and other formal pursuits that became possible since the late 1940s. Formal logic was somewhat buttressed by the growing prestige of science and technology. Although difficult and limited, exchanges with the outside world gradually expanded through translations, visits, and conferences. Conversely, the work of some Soviet philosophers, such as the semiotician Iurii Lotman (1922–1993) and his colleagues in the so-called "Moscow-Tartu School," found international recognition.

From the mid-1950s on some pre-Stalin figures reemerged. In Losev's prodigious output from 1953 to the time of his death in 1988 the partial truths of Marxism found their place among the broader principles of a phenomenologically modified Christian neoplatonism. Bakhtin's dialogic theories of culture, literature, and the (moral) self similarly rested on philosophical foundations that were sufficiently deep not to be perverted by adaptation to Soviet censorship. Unlike Losev who remained virtually unknown outside Russia, Bakhtin has become a towering presence in the western humanities.

The reappearance of these and other authors demonstrated that communism had not destroyed the continuity of the Russian intellectual tradition. This was largely due to Russian classical literature that remained even in the worst of times the backbone of all humanistic learning and education in Russia. The other key factor was the "Aesopian" writing, stemming from nineteenth-century polemics, by which philosophers masked (transparently enough for the reader to grasp) the true principles behind their critique of philosophy, art, religion, and culture. Losev delivered, for example a blistering critique of modernity in his *Estetika Vozrozhdeniia* (The aesthetics of the renaissance, 1978) that was tacitly based on an Eastern Orthodox view. A similar line of thought was pursued by younger philosophers such as Piama Gaidenko (b. 1934), Iurii Davydov (b. 1929), and Sergei Averintsev (1937–2004).

Characteristically, the most gifted among the newer generation of philosophers had to abandon classical Marxist materialism. Eval'd Il'entkov (1924–1979) and Merab Mamardashvili (1930–1990) exemplified opposing positions on the dialectical method, almost Hegelian in Il'entkov's case and almost openly neo-Kantian in Mamardashvili's. Yet another, mathematical-formalist, argument against the officially accepted dialectical materialism was developed by Aleksandr Zinov'ev (b. 1922),

who eventually had to emigrate and became a well-known writer.

In the 1970s and early 1980s censorship became more lax, allowing—although not without a struggle—the works of, and about, such authors as Solov'ev and Fedorov to be published. Alongside Vernadskii's ideas about the noosphere, Fedorov's doctrine of the "common cause" served as an inspiration for the loosely defined, nonofficial movement of cosmism. The latter was merged with Eurasianism by Lev Gumilev (1912–1992) who proposed a theory of ethnogenesis as a process affected by cosmic energy. All this signified a halting but perceptible expansion of the boundaries of philosophical discourse that increasingly weakened the hegemony of dogmatic Marxism.

Following Mikhail Gorbachev's reforms from the mid-1980s and until the dissolution of the USSR the hegemony of Marxism rapidly evaporated. One of the leading authors in official Marxism, Ivan Frolov (b. 1929) admitted in his study *Chelovek, nauka, gumanizm: novyi sintez* (Man, science, and humanism: A new synthesis, 1986) that the truths of Marxism were not, after all, absolute. The admission was an attempt to preserve the relevance of the doctrine in the new situation. The history of Soviet Marxism came to an end when the floodgates that held back previously suppressed philosophical literature, both Russian and foreign, finally opened. The return of the works of prerevolutionary and émigré Russian philosophers was the most remarkable part of this revival.

**RUSSIAN PHILOSOPHY ABROAD.** With the emigration after the 1917 Revolution and the expulsion of a large group of thinkers in 1922 Russian philosophy split into two strikingly unequal branches: the one inside and the other outside the country. The Bolshevik government's intolerance proved to be a blessing in disguise. While all independent philosophical thought was brutally suppressed in the Soviet Union, many of Russian philosophers abroad created the largest and the best part of their oeuvres. This was true of Berdyaev, Frank, Bulgakov, Shestov, and Il'in, as well as of the younger generation of philosophers among whom Georgii Florovskii (1893–1979) and Karsavin deserve special note. Russian thinkers in exile collectively created a body of literature that fulfilled the promise of the Silver Age as the Russian "religious-philosophical renaissance." A comprehensive evaluation of this literature remains a task for the future.

Among the diverse trends that existed in Russian philosophy abroad two seem particularly notable from

today's point of view: religious-philosophical and Eurasianist. The first was the continuation of the prerevolutionary religious idealism, whereas the second became yet another refraction of the old theme of Russia's destiny in a new situation created by the Bolshevik revolution. Berlin and then Paris were the centers of the first trend and Prague (as well as, briefly, Sofia), of the second.

Russian religious philosophy continued its preexile themes: critique of (Western) rationalism and the quest for integral knowledge; metaphysics of all-unity and sophiology; Russia's historical destiny cast in religious-idealist terms; and religious foundations of personhood.

The study of the history of Russian philosophy by this group became the culmination of the work begun in Russia. Zen'kovskii's two-volume *Istoriia russkoi filosofii* (History of Russian Philosophy, 1948–1950), Losskii's book of the same title (1951), Berdyaev's aforementioned essay on the Russian Idea, and Florovskii's *Puti russkogo bogosloviia* (Ways of Russian Theology, 1939) were towering achievements supplemented by numerous articles and essays by other authors. Their work was, collectively, the most important philosophical attempt to make sense of the Russian experience and especially of its last, vastly tragic phase. It is surprising how little would need to be changed, for example, in Frank's essay *Krushenie kumirov* (The Collapse of Idols, 1923), created before Stalin's repressions and World War II, if it were to be rewritten today.

Eurasianism began as a distinct movement with the publication of a collection titled *Iskhod k Vostoku* (Exodus to the East; Prague, 1921). It viewed Russia as straddling Europe and Asia in the geographic, geopolitical, and cultural-historical sense and enhanced the traditional Slavophile critique of the West by the Spenglerian sense of the "twilight" of Europe. Postcolonialist critique of Europe was prefigured in Eurasianism's claim that the western view of history merely promoted the West's ulterior interests under the guise of objective truth. The mistrust of the West was supplemented by the affirmation of the positive significance of the Asian element in Russian history and culture. Eurasianism had both a religious and a secular branch. The former was represented by such authors as Petr Savitskii (1894–1968) and Petr Suvchinskii (1892–1985), the latter by Florovskii and Karsavin. In Karsavin's case Eurasianism had a close affinity with the Solov'evian school. Nevertheless, for most Eurasianists religion was important only as a cultural-historical factor that contributed to the formation of Russia as a Eurasian entity. The religious theme in Eurasianism weakened especially after Florovskii left the movement. Some of his

secular opponents went so far as collaborating with the Bolshevik government that they saw as the heir to the cause of great Russian statehood. Those who returned to Russia, however, perished eventually in Stalin's concentration camps. Eurasianism as a political movement declined in the mid-1930s with the rise of National Socialism in Germany. Many of its members made significant contributions to the social and human sciences: George Vernadsky (1887–1973) in history, Nikolai Trubetskoi (1890–1938) and Roman Jakobson (1896–1982) in linguistics. Suvchinskii was a prominent musical critic. The political influence of Eurasianist ideas was restored to life in the post-Soviet period when they became a source of inspiration for a widely divergent spectrum of ideological schools of thought, ranging from nationalists dreaming of a new Russian Empire to Soviet-style Communists.

**POST-SOVIET PERIOD.** Rather than being resolved, philosophical questions were merely suspended by the ideological freeze during the Soviet period and once constraints fell old divisions quickly reemerged. During the early and mid-1990s Russian philosophers were primarily occupied with bringing back formerly suppressed patriarchy and rejoining the international philosophical dialogue. Berdyaev, Bulgakov, and Florenskii's writings were particularly favored during this period. But the list was quickly expanded to include the entire galaxy of Silver Age thinkers.

The second tendency—that is, restoration of contacts with the outside world—has by now resulted in a full spectrum of western and nonwestern influences without any apparent restrictions. Like several times earlier in history, Russian philosophers eagerly acquaint themselves with foreign philosophy: phenomenology, analytic philosophy, psychoanalysis, critical theory, poststructuralist thought, and a variety of nonwestern wisdom traditions. The old controversy between Slavophiles and Westernists was also apparently merely suppressed and has again become a notable factor in Russians' debates about their past, present, and future. The theme of the Russian Idea has returned in all of its prior permutations and now has been co-opted, among others, by communist authors who try to breathe new life into a doctrine that has lost much of its appeal. While the tradition of nonacademic philosophizing remains strong, the academy is now the backbone of philosophical life in Russia.

Soviet institutions, such as the Institute of Philosophy of the Russian Academy of Sciences and the journal *Voprosy filosofii* (Questions of Philosophy), have survived

their original ideological functions. The teaching of philosophy in higher education occupies the same place as elsewhere in the world and occurs without any ideological constraints. Literature for instruction in philosophy figures prominently among philosophical publications. The current output of academic philosophers embraces all disciplines of philosophy and represents all shades of opinion one finds elsewhere. Numerous works are published on the history of Russian philosophy; they include both special studies and historical surveys. Another notable feature is the striking decentralization of philosophical life that is no longer confined to "the capitals" but is active in many centers of higher learning in the country. There are no overwhelming political parties among Russian thinkers of the early twenty-first century. Neither the surviving communism nor the revived nationalism seem to hold commanding heights. If there is a threat to philosophy today it comes not from the state or radical ideology but from different quarters. Russian philosophy has joined contemporary western and nonwestern philosophical traditions in surviving the onslaught of mass culture. The new freedom and the rich intellectual, artistic, and literary legacy encourage hope, however, that Russian philosophy will rediscover not only its roots, but also the creative inspiration from which it first sprang.

**See also** Bakhtin, Mikhail Mikhailovich; Bakhtin Circle, The; Bakunin, Mikhail Aleksandrovich; Belinskii, Visarion Grigor'evich; Berdyaev, Nikolai Aleksandrovich; Bulgakov, Sergei Nikolaevich; Chernyshevskii, Nikolai Aleksandrovich; Chicherin, Boris Nikolaevich; Dostoevsky, Fëdor Mikhailovich; Fëdorov, Nikolai Fëdorovich; Florenskii, Pavel Aleksandrovich; Florovskii, Georgii Vasil'evich; Frank, Semën Liudvigovich; Herzen, Alexander Ivanovich; Ivanov, Viacheslav Ivanovich; Karsavin, Lev Platonovich; Kavelin, Konstantin Dmitrievich; Khomiakov, Aleksei Stepanovich; Kireevskii, Ivan Vasil'evich; Kozlov, Aleksei Aleksandrovich; Kropotkin, Pëtr Alekseevich; Lavrov, Pëtr Lavrovich; Lenin, Vladimir Il'ich; Leont'ev, Konstantin Nikolaevich; Losev, Aleksei Fëdorovich; Losskii, Nikolai Onufrievich; Lotman, Iurii Mikhailovich; Lunacharskii, Anatolii Vasil'evich; Mamardashvili, Merab Konstantinovich; Mikhailovskii, Nikolai Konstantinovich; Pisarev, Dmitri Ivanovich; Plekhanov, Georgii Valentinovich; Rozanov, Vasilii Vasil'evich; Shestov, Lev Isaakovich; Shpet, Gustav Gustavovich; Skovoroda, Grigorii Savvich; Solov'ëv (Solovyov), Vladimir Sergeevich; Tolstoy, Lev Nikolaevich; Trubetskoi, Evgenii Nikolaevich; Trubetskoi,

Nikolai Sergeevich; Trubetskoi, Sergei Nikolaevich; Zen'kovskii, Vasilii Vasil'evich.

## Bibliography

### PRIMARY SOURCES

Edie, J. M., J. P. Scanlan, and M. B. Zeldin, eds. *Russian Philosophy*. 3 vols. Knoxville: University of Tennessee Press, 1976.

Leatherbarrow, W. J., and D. C. Offord, eds. *A Documentary History of Russian Thought: From the Enlightenment to Marxism*. Ann Arbor, MI: Ardis, 1987.

*Russian Studies in Philosophy* (1962 to 1992 *Soviet Studies in Philosophy*). A quarterly translation journal.

### HISTORIES

Berdyaev, N. A. *The Russian Idea*. Translated by M. French. Hudson, NY: Lindisfarne Press, 1992.

Copleston, F. *Philosophy in Russia: From Herzen to Lenin and Berdyaev*. Notre Dame, IN: Notre Dame University Press, 1986.

Florovsky, G. *Ways of Russian Theology*. Translated by Robert Nichols. Belmont, MA: Nordland, 1979.

Kuvakin, V., ed. *A History of Russian Philosophy: From the Tenth to the Twentieth Centuries*. Buffalo: Prometheus Books, 1994.

Lossky, N. O. *History of Russian Philosophy*. New York: International Universities Press, 1951.

Masaryk, T. G. *The Spirit of Russia: Studies in History, Literature and Philosophy*. 3 vols. Translated by E. Paul and C. Paul, with additional chapters and bibliographies by J. Slavik. London: Allen and Unwin, 1955.

Scanlan, J. P., ed. *Russian Thought after Communism: The Recovery of a Philosophical Heritage*. Armonk, NY: M. E. Sharpe, 1994.

Walicki, A. *A History of Russian Thought from the Enlightenment to Marxism*. Translated by H. Andrews-Rusiecka. Stanford, CA: Stanford University Press, 1979.

Zen'kovskii, V. V. *A History of Russian Philosophy*. Translated by George Kline. New York, Columbia University Press, 1953; London: Routledge, 2003.

### REFERENCE LITERATURE

Alekseev, P. V., ed. *Filosofy Rossii XIX–XX stoletii: biografii, idei* [The philosophers of Russia in the 19th–20th centuries: Biographies, ideas]. Moscow: Akademicheskii Proekt, 1999.

de Lazari, A., ed. *Idee w Rosji. Idei v Rossii. Ideas in Russia*. Warsaw: Semper, 1999.

Maslin, M., ed. *Russkaia filozofia: Slovar'* [Russian philosophy: A dictionary]. Moscow: Terra, 1999.

*Vladimir L. Marchenkov (2005)*

## RUYSBROECK, JAN VAN

(1293–1381)

Jan van Ruysbroeck, the Flemish mystic, was born in the village of Ruysbroeck, near Brussels. He stood in close

relation to German contemplatives of the period, notably Meister Eckhart. In 1343 Ruysbroeck, together with two others, established a community at Groenendael that ultimately came under Augustinian rule. He was the prior of this community.

Ruysbroeck was not a trained theologian and had an imperfect knowledge of Latin. Though he made use in his mystical writings of language drawn from Eckhart, such as the “birth of Christ in the soul” and the “eternal Now,” he was sensitive to the kind of allegations of pantheism encountered by Eckhart and in fact directed against Ruysbroeck by Jean de Gerson. In his later writings in particular Ruysbroeck made it clear that he did not believe in the identification of the soul with God in the mystical state, and he criticized those contemplatives who gave up the active life and lapsed into quietism. He thus evolved a practical account of contemplation that connected it with good works.

Ruysbroeck distinguished between different phases of the good life, which should be practiced together. First, there is the active life of doing good works. This by itself will not bring blessedness, since it can mean moral self-reliance rather than dependence on God’s grace. But good works are a necessary part of the purification of the soul. Second, there is the practice of the inner virtues—faith, hope, and love. Third, there is the contemplative life, through which the soul may gain union with God. Those who attain this last condition are called “God-seeing.” They are not continually immersed, as it were, in this inner blessedness, but find themselves impelled to practice love and good works as a result of it. The practice of good works, suffused by the knowledge of God gained in the state of contemplative union, is what Ruysbroeck referred to as “the common life.” This, the ideal he tried to realize in his own monastic community, was interpreted as a reflection of the life of the Trinity, which was united in a common fruition analogous to that enjoyed by the mystic but was also outward-going through the creative power of God, analogous to the work of the monk in serving the society around him.

In order to illustrate the relation of union, yet difference, between the soul and God, Ruysbroeck made use of analogies drawn from human love, as the title of his major work, *The Adornment of the Spiritual Marriage*, indicates. Thus one should “rest in Him whom one enjoys.... There love has fallen in love with the lover, and each is all to the other, in possession and in rest” (*The Sparkling Stone*, 13). The love analogy had a certain aptness in bringing out both the sense of union and the necessary theistic distinction between the soul as creature

and the Creator. Ruysbroeck also made use of the Neoplatonic doctrine of eternal archetypes or forms, existing in God. Thus the ground of the soul is man's eternal archetype, and in realizing it in its purity and nakedness, the contemplative finds union with God. In this, Ruysbroeck, like other mystics of the period, exhibited the influence of Pseudo-Dionysius. He thus made use too of the notion that creatures proceed from God through the process of creation and return to him through contemplation. But since the creature needs to reflect the love displayed by God in the work of creation, so likewise the mystic must combine his return to God with the outgoing work of love.

Ruysbroeck's works were closely studied by those who belonged to the movement known as the Brethren of the Common Life, started in the latter part of the fourteenth century by Gerhard Groot, who knew Ruysbroeck. Thomas à Kempis belonged to this confraternity. Despite contemporary criticisms of his language as not always squaring with orthodox theology, Ruysbroeck was beatified by the Roman Catholic Church.

*See also* Augustinianism; Eckhart, Meister; Gerson, Jean de; Mysticism, History of; Pseudo-Dionysius; Thomas à Kempis.

### **Bibliography**

The definitive edition is B. Ponkens and others, eds., *Ruysbroeck Werken*, 2nd ed., 4 vols. (Cologne, 1950). English translations include a one-volume edition of *The Adornment of the Spiritual Marriage, The Sparkling Stone, and The Book of Supreme Truth*, translated by C. A. Wynschenk Dom and edited with an introduction and notes by Evelyn Underhill (London: Watkins, 1951), and *The Spiritual Espousals*, translated by E. Colledge (London: Faber and Faber, 1952). See also Ray C. Petry, *Late Medieval Mysticism* (Philadelphia: Westminster Press, 1957).

*Ninian Smart (1967)*

## RYLE, GILBERT

(1900–1976)

Gilbert Ryle, the British philosopher, was born in Brighton. Having read Classical Honour Moderations and the Final School of Literae Humaniores (Greats) he went on to read the then newly established School of Philosophy, Politics and Economics at the Queen's College, Oxford. He became a lecturer at Christ Church in 1924 and in the following year a student and tutor, and he remained there until his appointment as professor at the

end of World War II. He was the Waynflete professor of metaphysical philosophy in the University of Oxford from 1945 to 1968. Ryle was largely responsible for the institution of the new degree of bachelor of philosophy at Oxford. He served as the editor of *Mind*, after the retirement of G. E. Moore, from 1947 until 1971.

Ryle's philosophical writings covered a wide range of topics. They fall mainly within the fields of philosophical methodology, philosophical logic, and the philosophy of mind, but the total spread is very wide and includes some work on the history of philosophy, especially on Plato. Only the fields of moral philosophy, political philosophy, and aesthetics are comparatively neglected. Much of his writing takes the form of articles addressed to the solution of quite specific issues, and it is impossible to discuss here seriatim his "Negation," "Plato's *Parmenides*," "Conscience and Moral Conviction," and "Heterologicality," to mention the titles of only four such papers.

Probably the best approach to Ryle's philosophical work is through his views on the nature and method of philosophy, which have developed in a consistent way after the end of a short and early flirtation with phenomenology. Many of his articles on specific topics seem to have a clear subordinate aim of illuminating these questions, while such important writings as "Systematically Misleading Expressions," his inaugural lecture, *Philosophical Arguments*, and the book *Dilemmas* are explicitly devoted to them. That *The Concept of Mind* can be regarded as an illustration of his views on philosophical method is a tribute to the consistency of his theory with his practice, though it would be an injustice to treat it merely as such.

Ryle's well-known article "Systematically Misleading Expressions" is important as being easily the first, although incompletely worked out, version of a view of philosophy closely akin to that which Ludwig Wittgenstein was then beginning to work out independently, and which is often spoken of as having been first suggested by Wittgenstein. This view treats philosophy as the activity of removing fundamental conceptual confusions that have their source in our overreadiness to construe grammatical similarities and differences as indicative of logical similarities and differences. For example, since either unpunctuality or the unpunctual Smith may, with grammatical similarity, be said to be reprehensible, some philosophers are inclined to conclude that similar things are being said of two objects, Smith and unpunctuality; hence, the world is thought to be populated by two kinds of objects, universals and particulars. Again, since "Mr. Baldwin is a statesman" is grammatically similar to "Mr.



Pickwick is a fiction,” philosophers have been tempted to suppose that the world contained fictions alongside of statesmen.

However, Ryle’s view is not fully worked out at this stage. Writing in a climate of opinion in which philosophy was widely regarded as the activity of analysis by which the true logical form of facts was explicitly displayed and the test of adequate language was taken to be a one-to-one correspondence with the form of facts, he did not entirely free himself from its influence. As a result, he cannot regard the reformulation of statements in a way that removes misleading grammatical similarities as merely a useful expedient for making ourselves aware of important differences between them; the reformulation is still thought of as the revelation of the true form of the fact, so that “Baldwin is a statesman” is, in an absolute sense, a correct form of utterance, while “Pickwick is a fiction” is incorrectly formulated.

This anomalous relic of logical atomism caused Ryle uneasiness even then, and it does not appear again. If we neglect it, we may regard “Systematically Misleading Expressions” as an exposition of a view that Ryle never abandoned, although he did refine it. One such refinement is found in *Dilemmas*. Here it is claimed that many philosophical problems, if not all, immediately present themselves in the form of dilemmas: We find ourselves holding, without the possibility of sincere repudiation, two or more opinions that seem to be incompatible (that, for example, we often choose responsibly what to do, and that we are what we are through our natural endowment as modified by environment—the problem of free will). Such dilemmas must be overcome by showing that the apparent conflict is a consequence of conceptual confusion rather than by choosing one horn on which to be impaled.

The emphasis is somewhat different in *Philosophical Arguments*. While in “Systematically Misleading Expressions” and in *Dilemmas* the emphasis is on the activity of freeing ourselves from conceptual errors and puzzlement, in *Philosophical Arguments* the more constructive side of the procedure is stressed. By methodically determining what can and what cannot be said without absurdity, which inferences are valid and which are invalid, which grammatical parallels are likely to mislead and which are not, we come to see better the “logical geography” of our conceptual system—how different concepts are related to each other and what are the different roles that they play. There is no essential conflict between the view of Ryle’s philosophical procedures as “removing conceptual roadblocks” and “freeing conceptual traffic jams,” to echo the

metaphor employed in *Dilemmas*, and the more constructive view of them. Thus, it would be idle to ask whether, or at which stages, *The Concept of Mind* is correctly viewed as exposing the confusion of “the ghost in the machine,” into which we are led by grammatical analogies, or as mapping the extension and boundaries of such interrelated concepts as “will,” “intelligence,” “imagination,” “thought,” and the like; the two aspects are not thus separable.

Ryle often expressed this view of philosophy in terms of the notion of a category mistake, as in *The Concept of Mind*. A category mistake occurs when something is taken to belong to a different category from its true one. Neither in *The Concept of Mind* nor elsewhere is any serious attempt made by Ryle to give a rigorous account of the notion of a category, although there is a historical discussion of it in “Categories,” and Ryle sees this notion as akin to Bertrand Russell’s notion of type. Although this is a gap, it is probably of little direct importance to the argument of *The Concept of Mind*. The essential thesis here is that there is a special kind of confusion that can be illustrated by that of taking team spirit as an element in a game as being on equal footing with serving or receiving, of taking a division as a military formation as being on equal footing with its component regiments, of taking Oxford University as an institution as being on equal footing with its component colleges. Ryle then goes on to claim that traditional Cartesian dualism treats the mind as an entity on equal footing with the body and mental activities as being on equal footing with bodily activities, and that this is a confusion of the same kind as those in the three illustrative cases. The language of category mistakes is not essential; Ryle could have used his terminology of 1931 and said that just as the grammatical similarity of “Jones gave an exhibition of dribbling” and “Jones gave an exhibition of ball control” could mislead us into thinking that Jones was giving two independent and simultaneous exhibitions, so the grammatical similarities between our talk of mental and bodily activities could mislead us into thinking that they were independent and simultaneous activities.

Such a misconception Ryle calls the dogma of the ghost in the machine. He attempts to show its falsity in a series of chapters on the main aspects of mental life, in which the arguments fall into two main classes. On the one hand he tries to show that the dogma of the ghost in the machine fails in its explanatory task and is logically incoherent, leading to such logical evils as vicious infinite regresses. On the other hand he tries to show that a satisfactory positive account of mental phenomena can be

given, without invoking the ghost, in terms of such things as style of performance, dispositions to certain characteristic performances, and acquired skills. Thus, if a person does a physical action while thinking about what he is doing, we must take it not that the ghost discursively thinks and the bodily machine moves but that the person performs bodily in an appropriate way, while being disposed to perform other actions if the occasion arises.

One chapter in the book is a restatement of the argument of the paper "Knowing How and Knowing That," published in 1946, and it is a plausible inference that this paper was the germ from which the larger enterprise sprang. In that article Ryle suggested that philosophers commonly take it that knowing how to do something is knowing the truth of certain principles and applying them to an activity. He pointed out that although a given cook may learn to cook from a cookbook, the principles of cookery are logically a distillation from the practice of those who know how to cook, just as the principles of valid argument are a distillation from the practice of those who know how to argue. Thus, knowing how to do things, being able to perform intelligently, is logically independent of any interior theorizing; therefore it involves a display of intelligence that others can witness, rather than a mechanical event from which we have to infer a piece of unwitnessable ghostly theorizing. *The Concept of Mind* attempts to extend the same line of thought to other mental phenomena.

It should be noted that Ryle is not content with the "weaker" thesis that overt human actions must not be analyzed as mechanical events brought about by non-physical, ghostly activities. In fact, he adopts the far stronger thesis that all references to the mental must be understood in terms of, in principle, witnessable activities. We must not only avoid ascribing the skill of a skillful driver to a ghostly "inner" driver, but we must also explain all mental life, including emotion and feeling, in terms of the witnessable. Certainly it is this feature of his book that has led many, with considerable plausibility, to class Ryle as a philosophical behaviorist, though he repudiated this label in advance. Ryle, indeed, sometimes refers to "twinges," "throbs," "flutters," and "glows" in his characterization of feelings in a way hard to reconcile with behaviorism, but it is notoriously difficult to see how such terms are not a relic of the essentially private in Ryle's public world. By adopting this stronger thesis Ryle avoids well-known difficulties about knowledge of other minds and privacy; however, it is not clearly required for the basic program, and much that he has to say is independent of it.

Much of the interest of this modern classic is independent of the question whether Ryle succeeds in demonstrating any general thesis. The detailed discussions of thinking, knowledge, will, emotion, sensation, intellect, and the like have great independent interest. In the course of these discussions Ryle introduces a number of philosophical distinctions, such as those of "task and achievement," "avowal," and "mongrel-categorical," that have become the common tools of modern philosophical discussion. The whole character of philosophical discussion of the mind has been decisively changed, even in quarters where Ryle's conclusions are strongly challenged, by the appearance of *The Concept of Mind*.

Another set of problems to which Ryle devoted a number of papers are those concerned with the concept of meaning. Here his review of Rudolf Carnap's *Meaning and Necessity in Philosophy*, his "The Theory of Meaning," published in *British Philosophy in the Mid-century*, and his contribution to the symposium "Use, Usage and Meaning" in the *PAS* supplementary volume for 1961 deserve special mention. One main contention in these articles is that it is words that are the bearers of meaning, and whose meanings have to be taught and learned, rather than sentences. To learn a language is to acquire a vocabulary and a syntax; this language is then used in speech, which is an activity that one performs by means of a language. The sentence is a unit of speech, not of language. The theory of meaning is therefore concerned primarily with words, not with sentences; but this theory, Ryle holds, has been often vitiated by a simple model of meaning that he calls the "'Fido'-Fido" theory, one that seeks always to find as the meaning of a word something that stands to that word rather as the dog Fido stands to the name "Fido." J. S. Mill partly emancipated himself from the theory by distinguishing connotation from denotation, but he continued to say that meaning was connotation and denotation. In the review of Carnap mentioned before, Ryle attempts to show that the "'Fido'-Fido" theory is still not an outworn fallacy but something that continues to vitiate much sophisticated modern work.

It is notable that the bulk of Ryle's philosophical writing avoids, rather than lacks, any historical discussion. There is the very minimum of reference to even recent learned controversy, and the great philosophers are rarely given even a casual mention. In *The Concept of Mind* the expression "Cartesian dualism" is a nickname for a kind of view that Ryle had once held and to which he thinks many are prone rather than a genuine historical reference. However, this is a policy of segregating the his-

tory of philosophy from the treatment of problems, not a sign of lack of interest in the history of philosophy. Ryle's historical interests, though eclectic, are wide. They have, however, centered on Plato; in addition to already published articles on Plato, Ryle devoted much work to problems arising from the Platonic dialogues, and further publications in that field may be expected.

In conclusion, a word should be said about Ryle's highly individual style, for it is of more than literary interest. It is peculiarly his own, so that it would be impossible for anyone familiar with it not to recognize his work from even a few sentences. One hallmark is the freshness of the vocabulary; although he liberally coined technical terms when he needed them, he always avoided the well-worn counters of philosophical exchange. Another hallmark is that although the general style is informal, the choice of words is literary rather than colloquial; this is achieved by the use of a vocabulary more novelistic than learned. Although there is much close argument in his writing, the importance of the fresh language, the bold metaphor, and the terse epigram in giving the problem a striking presentation, in bringing down pretentious castles of learned jargon, and in making his own contention memorable is very great indeed.

**See also** Analysis, Philosophical; Artificial and Natural Languages; Behaviorism; Carnap, Rudolf; Categories; Meaning; Mill, John Stuart; Moore, George Edward; Language, Philosophy of; Philosophy of Mind; Plato; Propositions, Judgments, Sentences, and Statements; Russell, Bertrand Arthur William; Thinking; Type Theory; Wittgenstein, Ludwig Josef Johann.

## Bibliography

### WORKS BY RYLE

- Review of Martin Heidegger's *Sein und Zeit*. *Mind* 38 (1929).  
 "Negation." *PAS* supp. 9 (1929).  
 "Are There Propositions?" *PAS* 30 (1929–1930).  
 "Systematically Misleading Expressions." *PAS* 32 (1931–1932).  
 "Phenomenology." *PAS* supp. 11 (1932).  
 "Imaginary Objects." *PAS* supp. 12 (1933).  
*Locke on the Human Understanding*. Oxford: Oxford University Press, 1933.  
 "About." *Analysis* 1 (1933).  
 "Internal Relations." *PAS* supp. 14 (1935).  
 "Mr. Collingwood and the Ontological Argument." *Mind* 44 (1935).  
 "Unverifiability-by-Me." *Analysis* 4 (1936).  
 "Induction and Hypothesis." *PAS* supp. 16 (1937).  
 "Taking Sides in Philosophy." *Philosophy* 12 (1937).  
 "Categories." *PAS* 38 (1937–1938).  
 "Plato's *Parmenides*." *Mind* 48 (1939).

- "Conscience and Moral Convictions." *Analysis* 7 (1940).  
*Philosophical Arguments*. Oxford: Clarendon Press, 1945.  
 Inaugural lecture.  
 "Why Are the Calculuses of Logic and Mathematics Applicable to Reality?" *PAS* supp. 20 (1946).  
 Review of Marvin Farber's *Foundations of Phenomenology*. *Philosophy* 21 (1946).  
 "Knowing How and Knowing That." *PAS* 46 (1946–1947).  
 Review of Carnap's *Meaning and Necessity*. *Philosophy* 25 (1949).  
*The Concept of Mind*. London: Hutchinson, 1949.  
 "'If,' 'So' and 'Because.'" In *Philosophical Analysis*, edited by Max Black. Ithaca, NY: Cornell University Press, 1950.  
 "Logic and Professor Anderson." *Australasian Journal of Philosophy* 28 (1950).  
 "Thinking and Language." *PAS* supp. 25 (1951).  
 "Feelings." *Philosophical Quarterly* 1 (1951).  
 "Heterologicality." *Analysis* 11 (1951).  
 "The Verification Principle." *Revue internationale de philosophie* 5 (1951).  
 "Ludwig Wittgenstein." *Analysis* 12 (1951).  
 Contribution to *The Physical Basis of Mind*, edited by Peter Laslett. Oxford: Blackwell, 1950.  
 "Ordinary Language." *Philosophical Review* 62 (1953).  
 "Thinking." *Acta Philologica* 9 (1953).  
 "Pleasure." *PAS* supp. 28 (1954).  
 "Proofs in Philosophy." *Revue internationale de philosophie* 8 (1954).  
*Dilemmas*. Cambridge, U.K.: Cambridge University Press, 1954.  
 Introduction to *The Revolution in Philosophy*, edited by D. F. Pears. New York: St. Martin's Press, 1956.  
 "Sensation." In *Contemporary British Philosophy*, edited by H. D. Lewis, third series. London, 1956.  
 "The Theory of Meaning." In *British Philosophy in the Mid-century*, edited by C. A. Mace. London: Allen and Unwin, 1957.  
 "Predicting and Inferring." *Colston Papers* 9 (London, 1957).  
 "On Forgetting the Difference between Right and Wrong." In *Essays in Moral Philosophy*, edited by A. I. Melden. Seattle: University of Washington Press, 1958.  
 "A Puzzling Element in the Notion of Thinking." *Proceedings of the British Academy* 44 (1958).  
 Article on Wittgenstein in *Scientific American* (1959).  
 "Letters and Syllables in Plato." *Philosophical Review* 59 (1960).  
 Review of Richard Wollheim's *F. H. Bradley*. *Spectator* (1960).  
 "Comment on Mr. Achinstein's Paper." *Analysis* (1960).  
 "Use, Usage and Meaning." *PAS* supp. 35 (1961).  
 "La phénoménologie contre le *Concept of Mind*." In *La philosophie analytique*, edited by H. Bréra. Paris, 1962.  
 "Thinking Thoughts and Having Concepts." *Institut Internationale de Philosophie* (1962).  
 "Abstractions." *Dialogue* 1 (1962).

**J. O. Urmson (1967)**

## RYLE, GILBERT [ADDENDUM]

From Gilbert Ryle's death in 1976, and through the late 1990s, his views were not the focus of much philosophical attention. Studies of his thought have been published (e.g., Lyons 1980; Stroll 2001) and his character and role in Oxford have been illuminated by the memoirs of others (e.g., Mabbott 1986).

Ryle's own approach to the understanding of psychological concepts was superseded by the emergence of the psycho-physical identity theory, mainly because his analytical concentration on (behavioral) dispositions, of whatever complexity, seemed not to confer a sufficiently real status on lots of psychological processes, for example, feeling a pain or occurrent thinking.

His most lasting intellectual legacy has been the supposed distinction between knowing how and knowing that, which has remained part of philosophical folklore since he propounded it. According to Ryle knowing how to F is distinct from any knowledge that a proposition is true, amounting rather to a capacity to do the action in question. Ryle's distinction has been relied on by those (e.g., David Lewis) who have tried, when answering certain antimaterialist arguments, to give a practical, non-factualist account of knowing what an experience is like. However, considerable skepticism is being generated about Ryle's distinction (e.g., in Stanley and Williamson

2001; Snowdon 2003). Thus, someone who is injured can know how to do something even though he or she is unable to do it, and much knowhow seems to be knowledge that some way is the way to act. These criticisms have been resisted (e.g., Koethe 2002; Rumfit 2003). There is considerable debate and it remains to be seen whether this aspect of Ryle's thought will suffer the fate of the rest.

*See also* Behaviorism; Lewis, David.

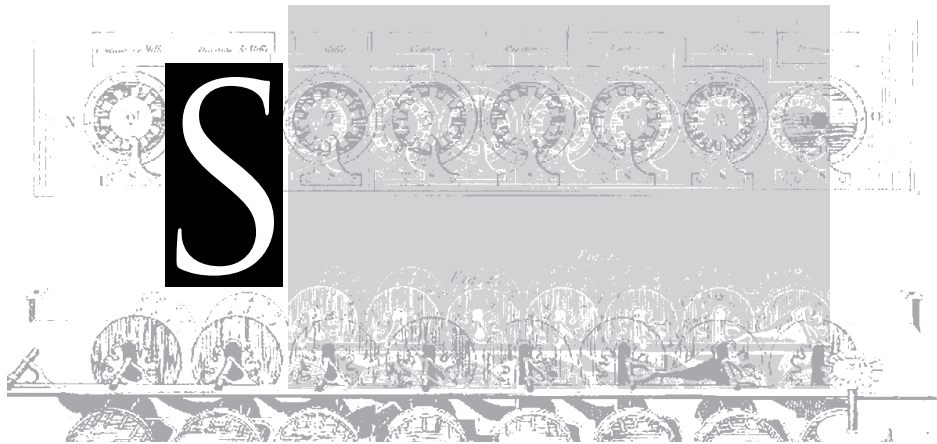
### **Bibliography**

- Koethe, J. "Stanley and Williamson on Knowing How." *Journal of Philosophy* 94 (2002): 325–328.
- Mabbott, J. D. *Oxford Memories*. Oxford: Thorntons, 1986.
- Rumfit, I. "Savoir Faire." *Journal of Philosophy* 100 (2003): 158–166.
- Snowdon, Paul F. "Knowing How and Knowing That: A Distinction Reconsidered." *Proceedings of the Aristotelian Society* 104 (1) (2003): 1–29.
- Stanley, Jason, and Tim Williamson. "Knowing How." *Journal of Philosophy* 98 (2001): 411–444.

### WORKS ON RYLE

- Lyons, William. *Gilbert Ryle: An Introduction to His Philosophy*. Brighton, U.K.: Harvester Press, 1980.
- Stroll, Avrum. "Gilbert Ryle." In *A Companion to Analytic Philosophy*, edited by A. P. Martinich and David Sosa. Malden, MA: Blackwell, 2001.

*Paul F. Snowdon (2005)*



## SAADYA

(882–942)

Saadya, sometimes called al-Fayyumi from the section of Upper Egypt in which he was born, had a brilliant career as the most distinguished intellectual leader of Jewry in his age. He was twenty-three when he left his Egyptian home to play his part on the wider stage of Palestine, Syria, and Babylonia. By this time he had already composed the first known Hebrew dictionary and an important treatise refuting the views of Anan ben David, the founder of the rationalistic Karaite sect. In 921, the rabbis of Babylonia challenged the authority of the Palestinian rabbis to fix the Hebrew calendar. Saadya's defense of the position of the Babylonian rabbis was most effective; he was rewarded by appointment to the rabbinical academy at Sura in Babylonia; and a few years later, in 928, he was the first non-Babylonian ever to be named as the head (*gaon*) of the academy. His tenure of this position was neither calm nor prolonged. Disputes with the exilarch of the Babylonian Jewish community led to the removal of Saadya and his retirement from active participation in the life of the community. His last years saw a burst of literary creativity.

The writings of Saadya truly signaled the birth of a new creative period in Jewish life. He was a pioneering student and productive scholar in many fields of Jewish concern, including Hebrew grammar and philology, biblical exegesis, and Jewish liturgy. The early attacks on the views of Anan were followed by a long series of writings against Anan's fellow sectarians; since Karaism, a movement that rejected rabbinical and Talmudic law, was at this time the major internal threat to the unity of Jewish life, Saadya's anti-Karaite polemics continued throughout his career. The primary activity of Saadya's public life was in the legal field, and here his contributions were outstanding. In addition to commentaries on Talmudic treatises, Saadya wrote at least ten systematic monographs on a variety of Jewish legal subjects; one of these, *Inheritance*, is preserved in its entirety in the Bodleian Library at Oxford. It was published in 1897 under the editorial care of Joel Mueller. Fragments of others still exist. Saadya was the first to translate the Old Testament into Arabic; this translation, still in use, is notable for its use of paraphrase where a literal translation would have been subject to censure for anthropomorphism. He also composed the earliest known commentary on *Sefer Yetzira* (The book of creation), an important work of the Jewish mystical tradition.

Thus his major philosophical work, *The Book of Beliefs and Opinions* (Arabic title, *Kitab al-‘amanat wali’ tikaḍat*; Hebrew title, *Sefer ha-emunoth weha-deoth*), probably completed in 933, is but one of a long list of eminent contributions for which Saadya is remembered. He was probably impelled toward a systematic consideration of the relation between the religious beliefs of Judaism and the opinions arrived at through rational investigation both by the comparable activities of Muslim philosophers—the *kalam* and other schools—and by the quasi-rational approach characteristic of most of the Karaite spokesmen. In the intellectual milieu of the tenth century, the philosophical issues with which Saadya was concerned were widely and thoughtfully debated. Muslim philosophers of this age had far more of the corpus of Greek philosophical literature available to them than had their compeers in the Christian West. Saadya’s *Book of Beliefs and Opinions* may best be described, therefore, as a philosophical apologetics for rabbinite Judaism. The Mu‘tazilite school of Muslim philosophers generally presented their systematic treatises in the form of theodicies, treating first of the unity of God and then of his justice. Saadya’s philosophical work is similarly patterned but assigns a rather larger share of the discussion to the second, ethical part than to the first, more purely metaphysical and theological one.

Prefaced to the ten sections into which the body of the work is divided is an introductory treatise in which Saadya justifies his engaging in this sort of philosophical enterprise. Here he enters into questions of the sources of human knowledge, the relations of belief and doubt, and the prevalent view that rational speculation necessarily leads to heresy. He argues that not the use of reason, but exclusive dependence on human reason is undesirable. Properly used, in combination with revelation, rational speculation supports revealed religion. From this discussion Saadya moves, in the first major section, to a proof of the doctrine of creation out of nothing and a refutation of twelve contrary views. In the second major section of his book, Saadya discusses the unity of God and demonstrates how the Christian doctrine of the Trinity is based upon a misinterpretation of certain scriptural verses. Treatise three defends the idea of a divine law for God’s creatures as a necessary demand of reason and urges the need for prophecy and prophets as the vehicle by means of which the divine law is transmitted to men.

From the fourth treatise to the end of the work, Saadya’s concern is more with ethical questions and the consequences for men’s future redemption of their obedience or disobedience to the divine precepts delivered by

the prophets. In these sections, he defends on rational grounds all of the major doctrines of the Jewish tradition. The tenth and last treatise is of slightly different character; it presents an ethic of the middle way as the proper guide to man’s conduct in the affairs of daily life. Thus we may say that Saadya concluded his work on religious philosophy with a secular ethic.

**See also** Jewish Philosophy; Mysticism, History of.

### **Bibliography**

*Les oeuvres complètes de Saadia* was edited by J. Derenbourg (6 vols., Paris, 1893–1896). Saadya’s *Book of Beliefs and Opinions* may be found in a translation from the Arabic by Samuel Rosenblatt in the Yale Judaica Series, Vol. I (New Haven, CT: Yale University Press, 1948).

For material on Saadya, see Jacob B. Agus, *The Evolution of Jewish Thought* (New York and London: Abelard-Schuman, 1959); Joseph L. Blau, *The Story of Jewish Philosophy* (New York: Random House, 1962); Julius Guttman, *Philosophies of Judaism*, translated by D. Silverman (New York: Holt, Rinehart and Winston, 1964); Isaac Husik, *History of Mediaeval Jewish Philosophy* (New York: Macmillan, 1916); Henry Malter, *Saadia Gaon: His Life and Works* (Philadelphia: Jewish Publication Society of America, 1921); and *Three Jewish Philosophers. Philo, Saadya Gaon, Jehuda Halevi*, selections with introductions by H. Lewy, A. Altman, and I. Heinemann, eds. (New York: Meridian Books, 1960).

**J. L. Blau (1967)**

## SAADYA [ADDENDUM]

882–942

Saadya’s intellectual career was characterized by polemic and defense of rabbinic Judaism. As head (*gaon*) of the rabbinic academy of Sura in Babylonia in the first third of the tenth century, Saadya felt compelled to respond to the Karaites, those who challenged the authority of rabbinic Judaism. Saadya’s major philosophical work, written in Arabic, *Kitāb al-Āmānāt wa’l-Ī-tiqādāt* (The book of doctrines and beliefs) should be understood in large part as a defense against the Karaites. Following the Muslim Mu‘tazilites, who emphasized divine justice and unity, Saadya’s *Āmānāt* focuses on creation, divine unity, divine law and justice, and reward and punishment in this world and posthumously.

Prefacing the particular discussions just noted, Saadya outlines the sources of human knowledge and understanding, and of error. The senses can be unreliable, reason may be derailed if inferential skills are lacking, and overarching all of this is the propensity to impatience in

inquiry. For Saadya, the sources of error and doubt are both intellectual and moral. Human beings are frail creatures, and as a result, out of benevolence God provided humankind with a superhuman source of insight and true belief—the prophetic tradition—encapsulated in scripture and in the oral tradition of the rabbis. In advance of people’s (slowly and laboriously) discovering the truth unaided, God has provided them the answers through his prophets. In this way revelation and reason, revealed religion, and rational speculation coincide over time. Revelation provides an anchor for humans as they quest for knowledge. As Alexander Altmann puts it on behalf of Saadya: “Revelation is not essentially superior, but historically prior to Reason and has an educational function in the evolution of humanity” (1946, p. 18 [2002]). Maimonides viewed Saadya’s project as less than philosophical, assuming conclusions that ought to be proved. But Saadya’s defense against this would be that revelation of the truth provides just the starting point for patient inquiry.

*See also* Jewish Philosophy.

### Bibliography

#### WORKS BY SAADYA

Altmann, Alexander, tr. *Saadya Gaon, The Book of Doctrines and Beliefs*. Abridged edition. Oxford: East and West Library, 1946. Reprinted with a new introduction by Dan H. Frank. Indianapolis: Hackett, 2002.

#### WORKS ON SAADYA AND THE BACKGROUND OF HIS THOUGHT

Altmann, Alexander. “Saadya’s Conception of the Law.” *Bulletin of the John Rylands Library* 28 (1944): 320–339.

Altmann, Alexander. “Saadya’s Theory of Revelation: Its Origin and Background.” In *Saadya Studies*, edited by E. I. J. Rosenthal, 4–25. Manchester, England: Manchester University Press, 1943.

Brody, Robert. *The Geonim of Babylonia and the Shaping of Medieval Jewish Culture*. New Haven, CT: Yale University Press, 1998.

Davidson, Herbert. *Proofs for Eternity: Creation and the Existence of God in Medieval Islamic and Jewish Philosophy*. 95–106. New York: Oxford University Press, 1987.

Davidson, Herbert. “Saadia’s List of Theories of the Soul.” In *Jewish Medieval and Renaissance Studies*, edited by Alexander Altmann, 75–94. Cambridge, MA: Harvard University Press, 1967.

Davidson, Herbert. “The Study of Philosophy as a Religious Obligation.” In *Religion in a Religious Age*, edited by S. D. Goitein, 53–68. Cambridge, MA: Association for Jewish Studies, 1974.

Goodman, Lenn E., tr. *The Book of Theodicy: Translation and Commentary on the Book of Job by Saadiah ben Joseph al-Fayyumi*. New Haven, CT: Yale University Press, 1988.

Goodman, Lenn. “Maimonides’ Responses to Saadya Gaon’s Theodicy and Their Islamic Background.” In *Studies in Islamic and Judaic Traditions II*, edited by W. M. Brinner and S. D. Ricks, 3–22. Atlanta: Scholars Press, 1989.

Goodman, Lenn. “Rational Law/Ritual Law.” In *A People Apart: Chosenness and Ritual in Jewish Philosophical Thought*, edited by D. H. Frank, 109–200. Albany: State University of New York Press, 1993.

Goodman, Lenn. “Saadiah Gaon’s Interpretive Technique in Translating the Book of Job.” In *Translation of Scripture Jewish Quarterly Review*, 47–76. Philadelphia, 1990.

Stroumsa, Sarah. “Saadya and Jewish *kalam*.” In *The Cambridge Companion to Medieval Jewish Philosophy*, edited by D. H. Frank and Oliver Leaman, 71–90. Cambridge, U.K.: Cambridge University Press, 2003.

Wolfson, Harry A. *The Philosophy of the Kalam*. Cambridge, MA: Harvard University Press, 1976.

Wolfson, Harry A. *Repercussions of the Kalam in Jewish Philosophy*. Cambridge, MA: Harvard University Press, 1979.

*Daniel H. Frank (2005)*

## SABATIER, AUGUSTE

(1839–1901)

Auguste Sabatier was perhaps the Protestant theologian most influential in the early twentieth century. Many Catholic modernists as well as Protestant liberals believed that his philosophy of religion had achieved its object, a reconciliation between the essential verities of Christian experience and the demands of science. Sabatier was a professor of reformed dogmatics at Strasbourg and Paris and a sometime journalist and literary critic. He ended his career as dean of the Theological Faculty of Paris.

Sabatier described his theory of religious knowledge as “critical symbolism.” By this he meant to indicate that religious doctrine and dogma are attempts to symbolize the primary and eternal religious experience (or consciousness) of the believer. He taught that the doctrines of historical religions are secondary, temporal, and transient symbols of this central religious experience. Christian dogmas, then, are necessarily inadequate attempts to “express the invisible by the visible, the eternal by the temporal, spiritual realities by sensible images.” Christ and his disciples through the ages have experienced the divine presence of God the loving Father and with it a sense of moral repentance and an inner energy of the spirit. As with all personal experience, no symbolic structure can act as substitute. Such structures are, in every field, merely hypothetical attempts to grasp experience.

Correspondingly, Sabatier held that the cosmologies, legends, dogmas, and statements about the world and man propagated by historical religions in an attempt to

express and communicate the fact of religious experience can claim only derivative and relative validity. Moreover, they are conditioned by the state of science and philosophy as understood by those who create such religious symbolism. And just as science and philosophy do not give absolute and final truth, neither does religious dogma—hence the decline of older religious symbolism with the progress of science. God lives in man’s consciousness, not in dogmas and cosmologies. Man’s need for and experience of God’s presence prove his existence. Science and philosophy are masters of their own proper domain. Thus, “God is the final reason of everything, but the scientific explanation of nothing.”

Sabatier’s critical symbolism was exceedingly Protestant in that it rejected Catholic dogmatic absolutism for the absolutism of justification by faith. It appealed to many modern religionists of his day because it seemed to retain valid science and yet avoid positivistic nihilism and agnostic defeatism. Putting personal experience above theories about experience, Sabatier’s approach was found congenial in an age that produced Henri Bergson and William James. Like them, Sabatier seemed to give moral claims and value judgments a renewed truth. To know a thing religiously, Sabatier held, is to experience the sovereignty of spirit and to estimate the object known as a means or obstacle to the true moral life of the spirit. Teleology is reintroduced along with objective value, and the meaning of life, as well as the will’s freedom to choose good or evil, is made manifest. Sabatier’s theories could easily be adapted to the neo-Kantian and neoidealist tendencies at work in philosophy, social science, political ideology, literature, and art in the new century. His continued influence seems assured, for by basing the truth of religion on the personal experience of the believer, he joined the long line of “crisis” and existential theologians of our time.

**See also** Bergson, Henri; James, William; Life, Meaning and Value of; Neo-Kantianism; Teleology.

### Bibliography

#### WORKS BY SABATIER

- Esquisse d'une philosophie de la religion*. Paris: Fischbacher, 1897. Translated by T. A. Seed as *Outlines of a Philosophy of Religion*. New York: Harper, 1957.
- Essai d'une théorie critique de la connaissance religieuse*. Paris, 1899.
- Les religions d'autorité et la religion de l'esprit*. Paris, 1903. Translated by L. Seymour as *Religions of Authority and Religions of Culture*. London: McClure, Phillips, 1904.

#### WORKS ON SABATIER

- Michalescu, J. *Darlegung und Kritik der Religionsphilosophie Sabatiers*. Bern: Buchdruckerei Scheitlin, Spring 1903.
- Vienot, J. *Auguste Sabatier*. Paris, 1927. Contains the best available bibliography.

*John Weiss (1967)*

## SAINT-HYACINTHE, THÉMISEUL DE (1684–1746)

The real name of Thémiseul de Saint-Hyacinthe, the French freethinker, was Hyacinthe Cordonnier. Born at Orléans, he was unjustly reported to be the son of Jacques Bénigne Bossuet. His ambitious mother induced him to change his name and to become a cavalry officer. Later he devoted himself to the study of ancient and modern languages in Holland, from which he had to flee because of a jealous husband and to which he later returned because he had seduced one of his pupils. He became an editor of the new *Journal littéraire* (1713) and wrote in favor of the moderns. In 1714 his anonymous *Le chef-d'oeuvre d'un inconnu*, a satire of pedantry, won him notoriety. He eloped to London in 1722 with the daughter of a nobleman. He stayed there for twelve years, became a member of the Royal Society, and began a long and gratuitous quarrel with Voltaire, whom he offended in a satirical play (*Déification d'Aristarchus Masso*, 1732). He returned to Paris in 1734 and later moved to Holland, where he died in 1746.

Three of Saint-Hyacinthe’s writings are worthy of mention. The first book, *Le chef-d'oeuvre d'un inconnu*, is a bizarre work that could easily be a satire on the *explication de texte* method, as it is practiced in some milieus. His last book, *Recherches philosophiques sur la nécessité de s'assurer par soi-même de la vérité* (1743), is a defense of the power of reason to find truth and of its right to do so. He also argues for the moral-sense theory, with which he probably became familiar during his stay in England. His discussion of words as signs of ideas points toward linguistic analysis. Other chapters deal with demonstration and evidence, matter and the soul.

In between these two works, Saint-Hyacinthe wrote his interesting *Lettres écrites de la campagne* (1721). This potpourri is a long conversation treating of many subjects, moral and epistemological. He discusses truth in the light of John Locke’s definition; evidence for certitude, following the Cartesian *cogito* and the principle of contradiction. He proposes a methodology for discovering



the truth that is also Cartesian. Most interesting is his recognition of the nihilistic challenge to moral values that was becoming more vigorous at the time. The longest section of the book expounds the argument that moral nihilism is justified and that all moral values disappear if God does not exist. Saint-Hyacinthe's real purpose was to urge men to believe in God, but the effect of his argument was more likely to lead them to immoralism, for he expounds that doctrine forcefully and endeavors to make it an impregnable position except in the face of God's existence. These little-known pages are notable as the most systematic exposition of moral nihilism before the Marquis de Sade. The *Lettres* had some success, and were translated under the title *Letters Giving an Account of Several Conversations Upon Important and Entertaining Subjects* (2 vols., London, 1731).

Among Saint-Hyacinthe's other publications are the *Lettres à Mme. Dacier* (1715, concerning the *querelle d'Homère*); "Lettre à un ami, touchant le progrès du déisme en Angleterre" (in his edition of *Mémoires concernant la théologie et la morale*, 1732); and the novel *Histoire du prince Titi* (1735).

L. G. Crocker (1967)

## SAINT LOUIS SCHOOL, THE

See Harris, William Torrey

## SAINT-SIMON, CLAUDE- HENRI DE ROUVROY, COMTE DE (1760–1825)

The French social philosopher, Claude-Henri de Rouvroy, Comte de Saint-Simon, the founder of French socialism, was the eldest son of an impoverished nobleman. He was educated privately by tutors, among them the encyclopedist Jean Le Rond d'Alembert. Beginning a military career at the age of seventeen, he took part in the American Revolution and was wounded at the naval battle of Saintes in 1782. Despite subsequent disclaimers, Saint-Simon actively supported some of the measures introduced by the French Revolution of 1789. He renounced his title; he also drew up the *cahier* of his locality for the Estates General and presided at the meeting at which his commune elected a mayor. Although his revolutionary zeal earned

him two certificates of *civisme*, his activities were not wholly disinterested. He took advantage of the sale at low prices of church and *émigré* property by making considerable purchases. He was arrested in 1793, but since it transpired that a mistake had been made, he was released the following year. He was active in political life under the Directory, among other things participating in the peace negotiations with the English at Lille.

Saint-Simon finally retired from governmental and financial activity and embarked on the career of writer and prophet that continued until the end of his life. He first studied physics for three years, at the same time forming friendships with a number of leading scientists and writers whom he helped to support. Later he traveled extensively, especially in Germany, England, and Switzerland. It was not until 1814, however, when he found an able and enthusiastic collaborator and disciple in the future historian Augustin Thierry, that his writings began to reach a wide public, particularly among the managers and businessmen who had risen to positions of influence during the Napoleonic era. The list of subscribers for his publication *L'industrie*, the first number of which appeared in 1816, included various prominent industrialists and bankers. The next year Saint-Simon's partnership with Thierry ended, and he began an association with Auguste Comte—an event of considerable significance, for it was in Comte's later work that some of Saint-Simon's fundamental conceptions were given more systematic and trenchant expression than their originator had been able to achieve. The collaboration between these two forceful personalities lasted for seven years but was finally broken by a quarrel in 1824, the year before Saint-Simon's death.

### IDEALS AND REALITY

"The philosophy of the eighteenth century has been critical and revolutionary; that of the nineteenth century will be inventive and destructive" (*Oeuvres complètes*, Vol. XV, p. 92). This remark accurately reflects the position that Saint-Simon envisaged himself as occupying in the history of political and social ideas. He in no way wished to underestimate the achievements of his Enlightenment predecessors the *philosophes*, who by their bold attacks upon the traditional frameworks of thought and their criticisms of existing institutions had prepared the way for the vast upheaval of the French Revolution. Saint-Simon saw in the writings of such men as Étienne Bonnot de Condillac and the Marquis de Condorcet anticipations of his own belief that human affairs should be approached in a scientific, Newtonian spirit of inquiry,

and he sympathized with their contention that religious dogmas had over the centuries become the means by which the mass of the people had been held in ignorant and superstitious servitude to their rulers. He also shared the humanitarian and internationalist ideals that had inspired the work of his predecessors. (His subscription to these ideals, apparent in all his main publications, was perhaps most distinctively expressed in *Nouveau Christianisme* [Paris, 1825], an essay that appeared at the very end of his life.)

On the other hand, Saint-Simon's work also pointed forward to the quite new ways of conceptualizing and interpreting social relations that were later to gain wide currency through the writings of Karl Marx. In particular, Saint-Simon exhibited a far firmer grasp of the conditions that determine and mold historical change than had earlier thinkers, and this profoundly affected the form taken by his own practical recommendations. Sincerely held utopian ideals, even when carefully worked out in detailed political programs, were by themselves quite useless, he held, if they did not take account of these conditions. Utopian changes, if put into effect, were likely to result in a vacuum that would eventually be filled by forces as undesirable as those which had been expelled. The destruction of outdated institutions was one thing; their replacement by others of lasting validity, adapted to the technological, economic, and social requirements of the time, was another. This was surely the lesson of the French Revolution. Had not the high hopes and aspirations that marked its beginning ultimately foundered in atrocities, suffering, and tyranny?

## HISTORICAL CHANGE

Despite the importance he assigned to it, Saint-Simon never set out his conception of historical change and development in a precise or systematic form. Like his other contributions to social theory, it was put forward in a somewhat disjointed and piecemeal fashion. Nevertheless, an outline of his view can be extracted from various works, notably from his writings in the periodical *L'organisateur* (Paris, 1819–1820). Saint-Simon spoke as if he had discovered a necessary law of evolution valid for all societies at all times, but the kernel of what he had to say was actually based upon a single instance, the transformation that had overtaken European society since the feudal period. The chief originality and importance of his analysis of how this change came about lay in his recognition of the role played by the emergence and conflict of classes and of the way in which such conflict issues in new forms of political organization and of ideology adapted

to the interests of the socially and economically dominant class. The institutions and beliefs of the Middle Ages fulfilled a perfectly intelligible, and indeed necessary, function from the point of view of the stage of development society had at that time reached (it is notable that Saint-Simon's approach to medieval history was considerably more sympathetic than that of either his Enlightenment predecessors or his liberal contemporaries).

Only later, with the enfranchisement of the communes, the emergence of a class of independent producers, and the subsequent growth of an industrial system of production under the impact of scientific and technological advances, did feudal organization become evidently anachronistic. Then the very features of the framework that had provided medieval society with the protection and unity of purpose it required impeded the free development of the new forces germinating within it. Thus, the seventeenth and eighteenth centuries witnessed the culmination of two major developments. On the one hand, there were increasingly effective attacks by the commons against privileges and institutions that had outgrown their social utility; on the other, the doctrines of the church, which during the Middle Ages had performed valuable services but which had been rendered obsolete by scientific discoveries, were subjected to a series of unanswerable criticisms. The net result was "the ruin of the old system in its parts and as a whole" (*Oeuvres complètes*, Vol. XX, p. 104).

## ECONOMIC AND POLITICAL PROGRAM

The lessons Saint-Simon drew from previous developments for his own time were far-reaching. Although the old order was in a general condition of dissolution, it had still not been wholly superseded. Many of the chief centers of power and influence remained in the hands of "more or less incapable bureaucrats" (*ibid.*, pp. 17–26), idlers, and ignoramuses who owed their positions to the accident of birth or inherited wealth and who were in effect no better than destructive parasites. To a considerable extent "men still allow themselves to be governed by violence and ruse" (*ibid.*). In order to remedy this state of affairs, Saint-Simon appealed directly to the leaders of the new class of *industriels*, claiming that the hour had arrived for them to take into their own hands the management of society and thereby complete the revolution that had been maturing for so long. Only if this were done could society be reorganized in a way that would ensure its direction by efficient administrators, men who would see that those who could make a genuinely pro-

ductive contribution to its advance and prosperity were no longer ignored or exploited and received, instead, their appropriate reward.

Yet despite his insistence on the need for social justice, Saint-Simon had little faith in political democracy. He envisaged a hierarchical system, characterized by equality of opportunity rather than equality of wealth and run on explicitly elitist lines. The central administration of the community would consist of three chambers—the chamber of invention, the chamber of examination, and the chamber of deputies. Of these the first was to consist of artists and engineers who would propose plans, the second of scientists who would critically assess the proposals and also control education, and the third of captains of industry whose function would be executive and who (Saint-Simon somewhat optimistically assumed) would give just consideration to the interests of all members of the industrial class, workers and managers alike. Saint-Simon appears to have thought that in the type of society he had in mind, which would be rationally planned in a manner advantageous to all, there would be little or no need for the use of force to compel obedience to law and that government in the traditional sense would no longer be required. There is a clear anticipation of the Marxian conception of the withering away of the state.

### ETHICS AND RELIGION

Saint-Simon was always conscious of the importance of moral and social ideals in helping to promote harmony and a sense of purpose in human communities. In medieval times the Christian religion had performed this role, and he thought that there was a place for a comparable system of beliefs, adapted to contemporary knowledge and interests, in any viable modern society. For the creation of such a system he initially looked to philosophy, but in his later years he recommended a return to the fundamental tenets of Christian teaching. The ethical doctrines of Christianity, he held, retained their validity even if the theological and metaphysical dogmas associated with them are no longer acceptable.

### INFLUENCE

It is impossible in a short space to do justice to the fertility and originality of Saint-Simon's thinking on what he called social physiology. An untidy, impatient, and inelegant expositor of his own ideas, he nonetheless understood the central issues of his time better than many of his contemporaries and exhibited a keener insight into the economic and technical realities that lie beneath the surface of political arrangements and change. Marx indis-

putably owed a significant debt to him, but Marx was only one among a host of nineteenth-century thinkers who profited in one way or another from Saint-Simon's perceptive and imaginative mind.

**See also** Alembert, Jean Le Rond d'; Comte, Auguste; Condillac, Étienne Bonnot de; Condorcet, Marquis de; Enlightenment; Marx, Karl; Social and Political Philosophy.

### Bibliography

#### WORKS BY SAINT-SIMON

- Oeuvres complètes de Saint-Simon et Enfantin*, 47 vols. Paris, 1865–1876.
- Lettres d'un habitant de Genève à ses contemporains*. Edited by A. Pereire. Paris, 1925. Written in 1803.
- Textes choisis*. Edited by J. Dautry. Paris, 1951.
- Selected Writings*. Translated, with an introduction, by F. M. H. Markham. Oxford: Blackwell, 1952.
- Henri Saint-Simon (1760–1825): Selected Writings on Science, Industry and Social Organisation*. Translated and edited by Keith Taylor. New York: Holmes and Meier, 1975.
- The Political Thought of Saint-Simon*. Edited by Ghita Ionescu. Oxford: Oxford University Press, 1976.

#### WORKS ON SAINT-SIMON

- Berlin, Isaiah. "Saint-Simon." In *Freedom and Its Betrayal: Six Enemies of Human Liberty*, edited by Henry Hardy. Princeton, NJ: Princeton University Press, 2002.
- Carlisle, R. B. *The Proffered Crown. Saint-Simonianism and the Doctrine of Hope*. Baltimore: Johns Hopkins University Press, 1987.
- Dondo, M. M. *The French Faust, Henri de Saint-Simon*. New York: Philosophical Library, 1955.
- Durkheim, Émile. *Le socialisme*. Edited by Marcel Mauss. Paris: Alcan, 1928. Translated by Charlotte Sattler as *Socialism and Saint-Simon*. Yellow Springs, OH: Antioch Press, 1958.
- Ferrarotti, Franco. "Claude-Henri de Saint-Simon." In *An Invitation to Classical Sociology: Meditations on Some Great Social Thinkers*. Lanham, MD: Lexington, 2003.
- Iggers, Georg G. *The Cult of Authority; The Political Philosophy of the Saint-Simonians*, 2nd ed. The Hague: Nijhoff, 1970.
- Manuel, F. E. *The New World of Henri de Saint-Simon*. Cambridge, MA: Harvard University Press, 1956.
- Plamenatz, J. P. *Man and Society*. London: Longmans, 1963. Vol. II, Ch. 2.

**Patrick Gardiner (1967)**

*Bibliography updated by Philip Reed (2005)*

## SAINT VICTOR, SCHOOL OF

The Augustinian house of canons at St. Victor in Paris was founded in 1108 by William of Champeaux, the cele-

brated logician and theologian who retired there from the schools of Paris after undergoing a religious conversion and after Peter Abelard's attacks on his realism. The abbey survived until the French Revolution, but in the twelfth and early thirteenth centuries it was especially famous for its public school and for the distinction of the masters and canons who resided and taught there. From William, St. Victor derived high religious ideals, a leaning toward the conservative theological tradition of the school of Anselm of Laon, and an active interest in the work of other Parisian schools. Its masters mediated between the theological orthodoxy and strictness of the Cistercians—Bernard of Clairvaux was a friend to St. Victor—and the intellectual adventurousness of such secular masters as Abelard. St. Victor in the twelfth century combined Scholasticism and mysticism and exerted a most powerful influence upon the development of both philosophical and theological thought in that century. Not only did it possess among its canons some of the ablest writers of the age but it also attracted as long-staying guests such celebrated teachers as Peter Lombard and Robert of Melun. Besides producing a wealth of literature, its leaders also contributed to the fall of Abelard, to the damping of the enthusiasms of the Chartrains, to the containment of Gilbert of Poitiers, and to the correction of the Christological errors that abounded in the mid-twelfth century.

### HUGH OF ST. VICTOR

St. Victor, unlike Chartres, was not devoted to the liberal arts. No commentary upon a nontheological text is known to have been written there, and purely literary writings were even relegated by the greatest Victorine, Hugh (d. 1141), to the position of mere appendices to the liberal arts. The Victorines did not encourage profane studies for their own sakes. The extreme, fanatical Walter (d. circa 1180) intemperately denounced the Aristotelian spirit of Abelard, Gilbert of Poitiers, William of Conches, and Peter Lombard. Absalon (d. 1203), too, warned against the dangers found in Aristotle.

Hugh vigorously challenged his humanist contemporaries who in the first half of the twelfth century thought more often of pagan philosophy than of Christ and his saints. Against the Chartrains he insisted upon the disparity between the cosmogony of Plato's *Timaeus* and Christian truth. Nonetheless, in his *Didascalicon* (which contains a program of Christian education) Hugh shows that he was thoroughly immersed in secular studies as the preliminary to divine science, for he considered the arts an indispensable aid to the understanding of

Scripture. Hugh sought to pass through knowledge to wisdom and to promote that participation in the divine Wisdom for which man was made. Similarly, Godfrey (d. after 1194) also affirmed that the liberal arts and theology were inseparable and that together they offered a complete education.

Philosophical elements are found scattered in the writings of the Victorines. Inheriting the Boethian-Aristotelian theory of abstraction, Hugh appreciated the necessity for logic without exalting it as highly as did Abelard. In physics Hugh maintained the atomic theory of matter and accepted the principle of the conservation of matter. His psychology was Augustinian, and he found the proof for the existence of the immaterial soul in the fact of its self-consciousness.

### RICHARD OF ST. VICTOR

Both Hugh and his disciple Richard (d. 1173) describe the ascent of the soul in contemplation; Richard especially is the theorist of the degrees of love. But whereas Hugh insisted upon the inadequacy of reason and the necessity for faith, Richard, who rivaled Hugh as a spiritual writer, was more scholastic and laid a stronger emphasis upon dialectic to supplement the traditional scriptural and patristic authorities. Inheriting from Anselm of Canterbury his zeal to search for the "necessary reasons" of faith and for an understanding of belief, he accounted for the trinity of persons in God in abstract style with a very original dialectic of mutual love; he was also the first medieval thinker to provide, in one of the great speculative achievements of the period, an empirical basis in the principle of causality for a proof of God's existence.

### VICTORINE THEOLOGY

Essentially the Victorines provided a theology for contemplatives within the cloister rather than for the schools. Hugh was a systematizer of theology on Augustinian lines, using dialectic when needed. Richard became the mystical doctor of the later Middle Ages. Both Hugh and Richard were biblical exegetes and spiritual writers, and it is for this that they and such other Victorines as Andrew (d. 1175) and Gamier (d. 1170) and the poet Adam were best known in the Middle Ages. Godfrey was more pronounced in his humanism, combining Chartrain Platonism and Aristotelian dialectic with Victorine spirituality. Achard (abbot 1155–1160) also mingled Augustinian theology with Chartrain Platonism, but all the Victorines concurred in wishing to turn knowledge into wisdom and the reader of the profane sciences into a contemplative. They always returned to the internal and external experi-

ences of the soul, and frequently to the use of allegory and symbolism in the penetration of divine truths. In the early thirteenth century the influence of Pseudo-Dionysius, which had been powerful upon Hugh, prevailed again upon Thomas Gallus, who was a forerunner of the mysticism of the later Middle Ages.

**See also** Abelard, Peter; Aristotle; Augustinianism; Bernard of Clairvaux, St.; Gilbert of Poitiers; Peter Lombard; Plato; Platonism and the Platonic Tradition; Pseudo-Dionysius; William of Champeaux; William of Conches.

### **Bibliography**

- Baron, R. *Science et sagesse chez Hugues de Saint-Victor*. Paris, 1957.
- Bonnard, Fourier. *Histoire de l'abbaye royale de Saint-Victor*. 2 vols. Paris, 1904.
- Delhay, P. *Le microcosmos de Godefroy de Saint-Victor*. 2 vols. Lille and Gembloux, 1951. Text and study.
- Dumeige, G. *Richard de Saint-Victor et l'idée chrétienne de l'amour*. Paris: Presses Universitaires de France, 1952.
- Ebner, J. *Die Erkenntnislehre Richards von St. Viktor*. Beiträge zur Geschichte der Philosophie des Mittelalters, Vol. XIX, no. 4. Münster in Westfalen: Aschendorff, 1917.
- Ethier, A. M. *Le de trinitate de Richard de Saint-Victor*. Paris: J. Vrin, 1939.
- Fritz, G. "Richard de Saint-Victor." In *Dictionnaire de théologie catholique*, Vol. XIII, Cols. 2676–2695. Paris, 1937.
- Geyer, B. *Die patristische und scholastische Philosophie*, 261–272. Basel, 1927.
- Godfrey of St. Victor. *Fons Philosophiae*, edited by P. Michaud-Quantin. Namur, Belgium, 1956.
- Hugh of St. Victor. *Didascalicon*. Translated by J. Taylor. New York: Columbia University Press, 1961.
- Klein, J. P. *The Theory of Knowledge of Hugh of St. Victor*. Washington, DC: Catholic University of America Press, 1944.
- Ostler, H. *Die Psychologie des Hugo von St. Victor*. Beiträge zur Geschichte der Philosophie des Mittelalters, Vol. VI, no. 1. Münster in Westfalen: Aschendorff, 1906.
- Richard of St. Victor. *De Trinitate*, edited by J. Ribailier. Paris: J. Vrin, 1958.
- Vernet, F. "Hugues de Saint-Victor." In *Dictionnaire de théologie catholique*, Vol. VII, Cols. 240–308. Paris, 1922.

*David Luscombe (1967)*

## SALMON, WESLEY

(1925–2001)

The American philosopher of science Wesley Charles Salmon was born August 9 in Detroit, Michigan, and died April 22 near Madison, Ohio. He pursued undergraduate studies at Wayne University and the University of

Chicago Divinity School, received an MA in philosophy from the University of Chicago in 1947, and a PhD in philosophy from the University of California at Los Angeles in 1950. His principal academic appointments were at Brown University, Indiana University, the University of Arizona, and the University of Pittsburgh; he retired from this last institution in 1999. At UCLA his dissertation advisor was the philosopher of science Hans Reichenbach and much of Salmon's subsequent work was influenced by Reichenbach's philosophy. A lifelong defender of empiricism, Salmon made significant contributions to a wide range of topics, primarily in explanation, causation, inductive inference, and the philosophy of probability.

### WORK

Beginning in 1971, Salmon developed a widely discussed alternative to Carl Hempel's covering law model of scientific explanation. The key element of Salmon's statistical relevance model was its insistence that explanatory factors must be statistically relevant to the occurrence of the event to be explained. This undermined in two ways Hempel's view that an explanation must lead people to expect the explanandum to occur. It showed that this condition was not necessary because events with low probability, such as the occurrence of lung cancer, can be explained in terms of statistically relevant factors such as cigarette smoking. It also showed that Hempel's model did not provide sufficient conditions for an explanation because irrelevant factors such as a man's taking birth control pills, when included in a Hempelian explanation, undermine the effectiveness of an explanation of his not getting pregnant.

In the course of developing the statistical relevance model, Salmon began to stress the importance of the causal relevance, rather than the statistical relevance, of explanatory factors; his 1984 book *Scientific Explanation and the Causal Structure of the World* contains an account of probabilistic causality grounded in an "at-at" theory of causation within which spatiotemporally continuous markable processes connecting cause and effect play a central role. The aim was to provide an account of causation free of appeals to counterfactuals and thus acceptable to an empiricist, yet different from Hume's in stressing the importance of connecting processes. In the light of criticisms that the markability criterion required tacit appeals to counterfactuals, Salmon abandoned it in the early 1990s and adopted a position where the transmission of conserved quantities was what distinguished causal from non-causal processes.

This appeal to causal processes and the distinction between genuine processes and pseudo-processes meshed with Salmon's interests in space-time theories. In that area he defended a causal theory of space and time within which the direction of time was to be grounded in causal asymmetries. He also maintained a long-term interest in conceptions of synchrony and in defending a conventionalist approach to simultaneity relations.

In his 1984 book, Salmon argued for a form of scientific realism based on the principle of the common cause. This principle states that if an association is observed between two types of event then, in the absence of a direct causal connection between instances of the events, there exists a common cause responsible for generating the association. This principle is general and can be used to argue for the existence of unobserved entities. It lies behind the reasoning used by Bertrand Russell in inferring the continued existence of a cat from its occasional observed appearances and was employed, Salmon claimed, by Jean Perrin in using the similarity of values obtained from different experimental techniques to determine Avogadro's number to argue for the reality of atoms. Despite its appeal, the principle does have its limitations. It is inapplicable in certain quantum mechanical situations where there are no hidden variables. It is also easy to find cases where two properties each increase over time but there is no common cause underlying the two. Nevertheless, Salmon's emphasis on this principle has led to an important new way of thinking about scientific realism.

Much of Salmon's early work concerned issues in probability and induction. For many years he defended Reichenbach's pragmatic vindication of induction, which argues that inductive inferences, more specifically the "straight rule" that projects the existing relative frequency of an event's occurrence into the future, is at least as likely to be successful as any other rule. In the light of criticisms by Ian Hacking, Salmon tempered his advocacy of this approach while continuing to insist on the importance of linguistic invariance for inductive rules.

Throughout his career Salmon defended a relative frequency interpretation of probability, including its use in his accounts of causation. For Salmon, the correct frequency to attribute to an event was the frequency within the broadest homogenous reference class to which the event belongs—that is, that class of events for which no further statistically relevant factors exist. Although he occasionally displayed sympathy for a propensity approach to probabilities and appreciated the role played by logical probabilities in Carnap's inductive logic, he

developed important criticisms of both. In many of his writings, Salmon argued that an objective form of Bayesian inference could illuminate a number of issues in the philosophy of science. Most notable, Salmon's insistence on preserving the context of discovery/context of justification distinction and on using Bayesian methods in the latter area led him to argue that Kuhn's account of theory choice could be made more objective by employing Bayesian techniques of theory justification.

In addition to his philosophical contributions, Salmon was an outstanding teacher and was much admired for his personal qualities. His introductory logic book was widely used as an undergraduate text, went into three editions, and was translated into five foreign languages. The exemplary clarity of his writing is evident in all of his publications.

### **Bibliography**

#### PRIMARY WORKS

- The Foundations of Scientific Inference*. Pittsburgh: University of Pittsburgh Press, 1967.
- Space, Time, and Motion*. 2nd ed. Minneapolis: University of Minnesota Press, 1981.
- Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press, 1984.
- Four Decades of Scientific Explanation*. Minneapolis: University of Minnesota Press, 1990.
- Causality and Explanation*. New York: Oxford University Press, 1998.
- Reality and Rationality*, edited by M. Salmon and P. Dowe. New York: Oxford University Press, 2005.

#### SECONDARY WORKS

- Fetzer, J., ed. *Probability and Causality: Essays in Honor of Wesley C. Salmon*. Dordrecht, Netherlands: D. Reidel, 1988. Contains a comprehensive annotated list of his publications through 1988.
- McLaughlin, R., ed. *What? Where? When? Why?: Essays on Induction, Space and Time, Explanation*. Dordrecht, Netherlands: D. Reidel, 1982. Contains a brief autobiographical essay.

*Paul Humphreys (2005)*

## SANCHES, FRANCISCO

(c. 1551–1623)

Francisco Sanches, a philosopher and physician, was born on the Spanish-Portuguese border, either in Tuy or Braga, of Marrano or New Christian parents. His family had moved to Portugal and then to southern France to escape religious and political persecution. The young Sanches studied at the Collège de Guyenne in Bordeaux, the same

school that his distant cousin, Michel Eyquem De Montaigne, attended. Sanches studied in Rome and then went to the University of Montpellier, where he received a degree in medicine in 1574. He was appointed professor of philosophy in 1585 and professor of medicine in 1612 at the University of Toulouse, where he had a successful career until his death in 1623.

One of Sanches's first philosophical writings that has survived is a letter to the Jesuit mathematician, Father Christopher Clavius, who had just edited Euclid's works and whom Sanches had met in Rome. Sanches offered a skeptical attack on the possibility of attaining genuine truth in mathematics. This was followed by his most famous writing, *Quod nihil scitur* (*That Nothing Is Known*). He soon thereafter wrote a critical examination of the astrological interpretations of the comet of 1577, *Carmen de Cometa*, published in 1578, and some commentaries on portions of Aristotle's writings, as well as many medical works. Sanches criticized various Renaissance naturalistic views, such as those of Girolamo Cardano, and may have actually debated Giordano Bruno in person in Toulouse.

In the letter to Christopher Clavius, Sanches attacked a form of the Platonic theory of knowledge. We cannot gain knowledge of things through mathematical study, because the objects studied by mathematics are not the natural, real ones encountered in human life. Rather, these objects are ideal, or maybe even impossible ones, such as points and lines. The mathematical relations that are demonstrated about such objects do not help explain anything in nature or experience, unless we happen to know independently that the experienced objects have mathematical properties, and also know that the principles of mathematics are in fact true. As far as we can tell, mathematics is just conjectural or hypothetical until we can independently determine the nature of things.

Sanches's *Quod nihil scitur* was written in 1576 and published in 1581. In it he develops his skepticism by means of a critique of Aristotelianism. He begins by asserting that he does not even know if he knows nothing. Then he proceeds to analyze the Aristotelian conception of knowledge to show why this is the case.

Every science begins with definitions, but definitions are nothing but names arbitrarily imposed upon things in a capricious manner, having no relation to the things named. The names keep changing, so that when we think we are saying something about the nature of things by means of combining words and definitions, we are just fooling ourselves. On the one hand, if the names assigned to an object such as man, such as "rational animal," all

mean the same thing, then they are superfluous and do not help to explain what the object is. On the other hand, if the names mean something different from the object, then they are not the names of the object. By means of such an analysis, Sanches worked out a thoroughgoing nominalism.

Sanches went on to examine the Aristotelian notion of science. Aristotle defines science as "disposition acquired through demonstration." But what does this mean? This is explaining the obscure by the more obscure. The particulars that one tries to explain by this science are clearer than the abstract ideas that are supposed to clarify them. The particular, Socrates, is better understood than something called "rational." Instead of dealing with the real particulars, these scientists argue about a vast number of abstract notions and fictions. "Do you call this science?" Sanches asked, and then replied, "I call it ignorance."

The method of Aristotelian science, demonstration, is next attacked. A demonstration is supposed to be a syllogism that produces science, but this involves a vicious circle rather than engendering any new information. To demonstrate that Socrates is mortal, one argues from "all men are mortal" and "Socrates is a man." The premises, however, are built up from the conclusion: the particular, Socrates, is needed to have a concept of man and mortality. The conclusion is clearer than the proof. Also, the syllogistic method is such that anything can be proven by starting with the right premises. It is a useless, artificial means, having nothing to do with the acquisition of knowledge.

Sanches concludes that science cannot be certitude acquired by definitions, neither can it be the study of causes, for if true knowledge is to know a thing in terms of its causes, one would never get to know anything. The search for its causes would go on ad infinitum as one studied the cause of the cause, and so on.

For Sanches, true science is the perfect knowledge of a thing—"SCIENTIA EST REI PERFECTA COGNITIO." Genuine knowledge is immediate, intuitive apprehension of all the real qualities of an object. Thus, science will deal with particulars, each somehow to be individually understood. Generalizations go beyond this level of scientific certainty, and introduce abstractions, chimeras, and so on. Sanches's scientific knowledge consists, in its perfect form, of experiential apprehension of each particular in and by itself.

Sanches showed that, strictly speaking, human beings were incapable of attaining certainty. The science

of objects known one by one cannot be achieved, partly because of the nature of objects and partly because of the nature of humankind. Things are all related to one another and cannot be known individually. There are an unlimited number of things, all different, so they could never all be known. And still worse, things change so that they are never in such a final or complete state that they can be truly known.

On the human side, Sanches devoted a great deal of time to presenting difficulties that prevent people from obtaining true knowledge. Our ideas depend on our senses, which only perceive the surface aspects of things, the accidents, and never the substances. From his medical information, Sanches was also able to point out how unreliable our sense experience is, how it changes as our state of health alters. The many imperfections and limitations, with which God has seen fit to leave us, prevent our senses and our other powers and faculties from ever attaining any true knowledge. The conclusion of all this is that the only truly meaningful scientific knowledge cannot be known. All that humans can achieve is limited, imperfect knowledge of some things that are present in their experience through observation and judgment.

Sanches's claim that *nihil scitur* is argued for on philosophical grounds, on a rejection of Aristotelianism and an epistemological analysis of what the object of knowledge and the knower are like. His totally negative conclusion is not the position of Pyrrhonian skepticism, the suspense of judgment as to whether anything can be known, but rather the negative dogmatism of the Academics. A theory of the nature of true knowledge is asserted, and then it is shown that such knowledge cannot be attained. The Pyrrhonists, with their more thoroughgoing skepticism, could neither assent to the positive theory of knowledge, nor to the definite conclusion that *nihil scitur*.

Sanches put forward a procedure, not to gain knowledge, but to deal constructively with human experience. This procedure, for which Sanches introduced the term, for the first time, of *scientific method*, "Método universal de las ciencias," consists in careful empirical research and cautious evaluation of observable data. In advancing this limited or constructive view of science, Sanches was the first Renaissance skeptic to conceive of science in its modern form, as the fruitful activity about the study of nature that remained after one had given up the search for absolutely certain knowledge of the nature of things.

Sanches was influential in his own day and throughout the seventeenth century. *Quod nihil scitur* was reissued several times up to 1665. Late in the seventeenth

century two refutations appeared in Germany. People have seen possible influences not only on Descartes, but also on Pierre Gassendi, Marin Mersenne, Spinoza, and Leibniz, among others, although it is hard to delineate his exact influence as different from that of Montaigne, Sextus Empiricus, Cicero, Charron and other available skeptical sources who were read by most intellectuals of the time. It may be that Sanches's formulation of the skeptical problem is closer to the modern idiom than that of any of his contemporaries, including Montaigne, and in terms of how philosophy developed, reads more like a precursor of Bacon or Descartes.

*See also* Scientific Method.

### Bibliography

#### WORKS BY SANCHES

- Que nada se sabe*. Buenos Aires: Editorial Nova, 1944. Spanish translation of *Quod Nihil Scitur*, with an introduction by Marcelino Menéndez y Pelayo.
- Opera Philosophica*, edited by Joaquim de Carvalho. Coimbra, Portugal: 1955. Introduction in Portuguese by Carvalho.
- Tratados filosoficos*. 2 vols. Lisbon, Portugal: Univ. de Coimbra, 1955. Latin-Portuguese edition.

#### WORKS ABOUT SANCHES

- Coralnik, A. "Zur Geschichte der Skepsis. I. Franciscus Sanchez." *Archiv für Geschichte der Philosophie* 27, Neue Folge vol. 20 (1914).
- Franck, Adolphe, ed. *Dictionnaire des sciences philosophiques*. Paris: Hachette, 1875. See "Sanchez, François."
- Iriate, J. *Kartesischer oder Sanchezischer Zweifel*. Bottrop: W. Postberg 1935.
- Limbrick, Elaine, and Douglas F. S. Thomson, eds. *That Nothing Is Known*. Cambridge, U.K.: Cambridge University Press, 1988.
- Menéndez y Pelayo, Marcelino. "De los orígenes del criticismo y del escepticismo y especialmente de los precursors españoles de Kant." In his *Ensayos de critica filosófica*. Madrid: V. Suárez, 1918.
- Moreau, Joseph. "Doubte et savoir chez Francisco Sanches." In *Aufsätze zur portugiesischen Kulturgeschichte*. 1st ser., vol. 1: *Portugiesische Forschungen der Görresgesellschaft*. Munster in Westfalen, Germany: 1960.
- Moreira de Sá, Artur. *Francisco Sanches filósofo e matematico*. 2 vols. Lisbon, Portugal: Oficinas graficas casas portuguesas 1947.
- Owen, John. *The Skeptics of the French Renaissance*. London: S. Sonnenschein, 1893.
- Popkin, Richard H. *The History of Skepticism from Savonarola to Bayle*. New York: Oxford University Press, 2003.
- Senchet, Émilien. *Essai sur la méthode de Francisco Sanchez*. Paris: V. Giard and E. Brière, 1904.

**Richard Popkin (1967, 2005)**



## SANCTIS, FRANCESCO DE

See *De Sanctis, Francesco*

## SANTAYANA, GEORGE

(1863–1952)

George Santayana, the philosopher and man of letters, was born in Madrid. His parents separated within a few years of his birth, and his mother went to live in Boston, Massachusetts, with the children of a previous marriage. Santayana grew up in Ávila under his father's care, but at the age of eight he joined his mother in Boston. He was educated at the Boston Latin School and at Harvard College. After graduating from Harvard in 1886, he studied in Germany for two years and then returned to take his doctorate at Harvard, for which he wrote a thesis on Rudolf Lotze. He subsequently joined the department of philosophy and remained a member of the Harvard faculty until 1912, when a small inheritance permitted him to retire. He lived in England for a number of years and then in Paris, but in 1925 he finally settled in Rome. During World War II, he took refuge in the convent of an order of English nuns in Rome, and he continued to live there until his death.

## CULTURAL BACKGROUND

Both Santayana's personal life and his philosophical development were decisively influenced by his peculiar position as a Spanish Catholic living and teaching in a predominantly Protestant society with a philosophical and cultural tradition that he felt to be in many respects deeply alien to his own personality. He was always proud—rather defiantly so—of his Catholicism and his Latinity, despite the fact that he was not a believer and was not notably attached to Spain or to Spanish culture. These loyalties expressed instead a deeply rooted hostility to the commercial and democratic ethos of modern industrial society and an equally deep aspiration toward a radically different style of life and thought that, for Santayana, was best exemplified in the classical Mediterranean world. Philosophically, he felt his truest affinities to be with the Greeks and perhaps the Hindus, and among the moderns, with Benedict de Spinoza, rather than with the empiricism and idealism of German and Anglo American philosophy. In fact, however, his points of affiliation with the European and American philosophy of the modern period are both numerous and obvious, and it would appear that his debt to the post-Cartesian tradition in modern philosophy is much

greater than he was inclined to think. What chiefly set his work apart from the mainstream of twentieth-century philosophy was his highly personal and literary mode of writing and his rather disdainful lack of interest in the methodological questions that were of central importance to the development of phenomenology on the Continent and analytic philosophy in the English-speaking world. When one considers the substantive doctrines to which he was committed, however, and, in particular, the ontological distinctions on which his "Realms of Being" rest, his philosophy emerges as a highly idiosyncratic doctrine of transcendental subjectivity that would scarcely be conceivable apart from the very tradition of modern philosophy which he so violently criticized.

## PHILOSOPHICAL DEVELOPMENT

Santayana's philosophical career falls naturally into two main periods. The first of these is the period in which he published *The Sense of Beauty* (1896) and *The Life of Reason* (1905–1906); its chief distinguishing feature is Santayana's disposition at that time to conceive of philosophy as a kind of descriptive psychology of the higher mental functions. He assumed the broad truth of the doctrine of biological evolution and its relevance to the understanding of mental phenomena, and while he held all knowledge to be representational in nature, he did not question "our knowledge of the external world," nor did he feel the need for any initial withdrawal of belief in such a world in the Cartesian manner. "Mind" is placed firmly in its biological context, and such independence as it enjoys is due not to any special ontological status, but rather to its capacity for giving an ideal and aesthetic meaning to its natural setting and functions.

In the second period, during which he wrote *Scepticism and Animal Faith* (1923) and *Realms of Being* (1927–1940), Santayana came to feel the need for a greater systematic rigor in the exposition of his views and for a purified and nonpsychological mode of stating the fundamental distinctions on which his philosophy rested. In particular, he felt that in *The Life of Reason* he had not made clear enough that the "nature" described there as having been "drawn like a sponge, heavy and dripping from the waters of sentience" was the idea of nature, not nature itself. He now tried to correct this error by means of a set of ontological—that is, nonpsychological—distinctions between the different kinds of being that are the objects of different kinds of mental activity. Thus, imagination, for example, must be defined by reference to the essences or abstract characters that Santayana now recognized as having a distinct ontological status, rather than

the other way around. In carrying out this revision of his earlier views, Santayana was in some measure aligning himself with similar antipsychologistic tendencies at work in the logical realism of Bertrand Russell, as well as in the phenomenology of Edmund Husserl, which he regarded as having a certain affinity to his own views.

Some commentators have felt that this shift from what they describe as Santayana's earlier naturalism to his later "Platonism" amounted to a fundamental change in his general philosophical perspective. Santayana's own statements, however, make it clear that the system presented in *Realms of Being* is to be understood as the ultimate philosophical basis of the naturalistic *Weltanschauung* sketched out in *The Life of Reason*, in which he had paid relatively little attention to technical philosophical issues. It must be admitted that the moral atmosphere of the two works differs, and that in the later one Santayana seems even more the detached spectator of the noncontemplative phases of the "life of reason" than he had before. But this is as much a personal as a philosophical matter, and there is no good reason for denying the fundamental unity of Santayana's thought during the two main periods of its development.

## AESTHETICS

Santayana's first important philosophical work was *The Sense of Beauty* (1896). In it he attempted to state a complete aesthetic theory, which he later developed further in *Reason in Art* (1905), Volume IV of *The Life of Reason*. In the earlier book, aesthetic theory is characterized as a psychological inquiry whose data are aesthetic judgments considered as "phenomena of mind and products of mental evolution"; the inquiry is to be distinguished both from the actual exercise of critical judgment and from the historical investigation of the evolution of the various art forms. Santayana argued that this inquiry must be carried out independently of metaphysical issues and the "interests of the moral consciousness," and that it must make clear the bases of aesthetic experience in human nature as conceived by natural science and in particular evolutionary biology. To this end, Santayana sketched out a theory of value according to which all preference is an essentially irrational expression of vital interest and the standard of value is the enjoyment or pleasure procurable through different courses of action. Morality is concerned with negative values, namely, the avoidance of pain and suffering, while aesthetic value is concerned with positive enjoyment and stands in the same relation to morality as play does to work.

The pleasure that is distinctively aesthetic, however, must be further qualified as intrinsic (or immediate) and as "objectified," in the sense of being experienced as a quality of a thing and not as an affection of the organ which apprehends it. Santayana denied that it must have the disinterested character attributed to it by Immanuel Kant and that it must be universally shared. He defined beauty as "pleasure objectified."

**MEDIUM, FORM, EXPRESSIVENESS.** Santayana added to this definition of beauty a threefold distinction between the materials of a work of art, its form, and its expressiveness. Of these, the first two are intrinsic features of the work of art, which thus consists of sensuous elements that have varying degrees of aesthetic value by themselves, and a form or arrangement by means of which these elements are unified and which has its own distinctive value. This synthesis, which constitutes form, is "an activity of the mind." While Santayana throws out suggestions as to how the nature of our perceptual apparatus may determine *which* forms give pleasure, these suggestions are never developed, and there is a heavily mentalistic cast to his whole account of aesthetic experience. This is particularly true of his treatment of expression, which is the power of a work of art to suggest images and ideas that, by becoming associated with it, enhance its value. These associated values may be aesthetic, practical, or moral; or they may be intellectual, as they are in the case of those forms of art, for example, tragedy, which present the ugly as well as the beautiful, and whose value thereby consists in satisfying our desire to know life as a whole. In the end, however, while these distinctions of materials, form, and expression have the validity proper to their spheres, the experience of beauty remains, according to Santayana, unique and unanalyzable.

**FUNCTION OF ART.** In *Reason in Art* Santayana was concerned with the place of art, as one good among many, within the moral economy of the life of reason. He distinguished between the practical arts and the fine arts and explained the emergence of the latter from the former through the gradual growth of an appreciation of the intrinsic value of what originally had merely instrumental value. Applying this principle, Santayana described the development of music, poetry, and the plastic arts, and in each case attempted to relate the special features of the artistic medium to the mode of abstraction and selectivity that is peculiar to a given art form. He treated all works of art as more or less abstract symbolizations of the natural environment and interests of human beings, and as being animated by an internal "dialectic" of their own

through which the moral and dramatic unities of our experience are indirectly expressed. There can be no absolute or universal principles for criticizing works of art, since our critical judgments are simply the corrections or modifications that our aesthetic preference undergoes in the wake of experience; and there is no a priori guarantee that these corrections must be convergent. The ultimate justification of art is simply that it adds greatly to human enjoyment, and thus to human happiness.

### THE LIFE OF REASON

Santayana intended *The Life of Reason; or The Phases of Human Progress* (1905–1906) as a naturalistic biography of the human mind, but as he himself pointed out, it was at least partially inspired by G. W. F. Hegel's *Phenomenology of Mind*. What appealed to Santayana in that work and similar ones in the idealistic tradition was the idea of sympathetically espousing the changing perspectives—scientific, moral, religious, and aesthetic—by which the mind progressively defines its relationship to its natural milieu. By beginning with *Reason in Common Sense*, he hoped to avoid the fundamental error of the idealists, which was to lose all sense of the dependency of this evolution upon a nonmental nature and of its responsiveness to the strains and stresses of our animal being. For the fraudulent dialectical necessity that Hegel had imposed on human history, Santayana proposed to substitute an appraisal—in the broad sense, a moral appraisal—of the contribution made by each of these phases of human development to the ideal of a rational and happy life.

**REASON AND IMAGINATION.** In *Reason in Common Sense*, the discovery of natural objects is described as the first and irreversible achievement of human reason operating upon the materials of sense experience. Knowledge of these objects is inevitably representative and indirect, and the relationship of thought to reality must be conceived as an ideal correspondence and not as a material appropriation. Coordinate with these “concretions in experience” are “concretions in discourse,” or concepts which sustain among one another all manner of “dialectical” relationships; and the active elaboration of these is the generic activity of imagination. Imagination becomes understanding when, almost by accident, some of its structures prove to be faithful transcriptions of a sequence of natural events; but even when the understanding is most successful, there remain unassimilable traits of experience which, at best, have a tangential relation to the natural order.

Toward the free creative activity of the imagination itself, Santayana maintained a dual attitude. It must not, he said, be allowed to impose itself as a literal rendering of what exists, as it all too often attempts to do. When it is allowed to do so, it can only produce a fantastic physics in which dramatic and moral unities are substituted for unities of fact and real process. In another sense, however, the life of reason is the life of the imagination, and its function of idealization and symbolic transformation yields the highest and purest enjoyments of the mental life. Even when the imagination becomes practical, as it does in science, it is the intrinsic aesthetic value of its creations, and not their ulterior practical use, which gives them a place within the life of reason. But at the same time that he praised the imagination, Santayana continually warned against the tendency to confer substantial reality upon the essences it elaborates and to assign to them a causal efficacy within the order of nature. The only power that Santayana was willing to attribute to consciousness itself was that of conferring meaning and ideal unity upon events, and it is in this sense that he described himself as being a materialist.

**RELIGION.** If Santayana's theory of the imagination finds its most natural application in his treatment of art, an area in which the claim to any literal validity is reduced to a minimum, the case of religion, which he considers in *Reason in Religion*, Vol. III of *The Life of Reason*, is somewhat different. Religion, Santayana said, is a poetic transformation of natural life in the interest of the moral ordering of that life, even though each religion is typically regarded by its followers as embodying a literal truth. Religion is myth, and it presents “an inverted image of things in which their moral effects are turned into their dramatic antecedents.” Because it is myth, religion must not be judged by the inappropriate standard of literal truth, but on the basis of the imaginative richness and comprehensiveness of its reorganization of our moral experience. One's religion is in fact something like one's language or nationality—a native idiom of the moral life which may have its imperfections, but which is both difficult and unwise wholly to abandon. Mystical religions are those that effect vast simplifications of the moral life by excluding all but one element in the natural life, while fanatical religions are those that suppress, on the authority of their own unique truth, all forms of moral poetry other than their own. In Santayana's view, both are inimical to the true value of religion, which is the encouragement it gives us to live in the imagination. True religion stimulates both piety, which Santayana defined as “man's reverent attachment to the sources of his being and the

steadying of his life by that attachment,” and spirituality, which liberates us from the harsh realities of animal need and desire by interposing an ideal meaning—one that assigns to the goods of this world their proper and subordinate place.

What is paradoxical in Santayana’s philosophy of religion is the fact that while he treated all religions as having, at best, a symbolic or expressive truth, he severely condemned the liberals and “modernists” who have attempted, while remaining within the church, to substitute for the literalistic dogmatism of the past a view of religion that in many respects resembles the one held by Santayana himself. It seems inconsistent to deny that a claim to literal truth is essential to religion and at the same time to require that those who surrender this claim must leave the church. This is perhaps a special case of a general paradox resulting from the fact that while Santayana declared “spirit” to be wholly inefficacious, it is an intrinsic feature of the life of reason that spirit should view itself as having efficient power. One may also speculate as to whether Santayana’s distaste for views resembling his own, when they become more than the private insights of detached and passive observers and are applied to the task of modifying some institution such as a church, did not itself express a social attitude and a partisanship that cannot claim any special philosophical justification.

**SOCIAL THEORY.** Santayana’s theory of society is stated in *Reason in Society*, Volume II of *The Life of Reason*, and also, in expanded form, in *Dominations and Powers* (1949), his last major work. In the main, social life is assigned a subordinate role within the life of reason. Its principal task and justification is the generation of, and care for, human beings, and it serves ideal ends only incidentally. Society originates in the reproductive instinct, and while this instinct lends itself readily to imaginative development, it finds its ultimate fruition in institutions (the family, the army, the state) that are predominantly practical in nature and, at best, capable of a retrospective idealization. It is, of course, possible for individuals to become associated with one another outside the disciplinary framework of these primary institutions, and when they do so freely, on the basis of a common allegiance to an ideal, they form what Santayana called a “free,” or “rational,” society. Patriotism is the loyalty they feel to such societies; but the deepest loyalties of the life of reason are not to anything actual, but to the ideal presences of which, Santayana said, our human partners in the pursuit of the ideal, as well as we ourselves, are at best imperfect symbols. Thus it turns out that the true society—the

only society that is a perfect instrument of the life of reason—is the society of the mind and of the essences it entertains.

If Santayana’s theory of society expresses, as indeed it does, a profound lack of interest in the practical concerns by which any human society is principally animated, he was nevertheless not without his own strong preferences with regard to a certain ordering of society. A pervasive animus against democracy and liberalism runs through all his discussions of society and is perhaps most noticeable in *Dominations and Powers* (1949). Human society, Santayana argued, is necessarily aristocratic and hierarchical, and egalitarian democracy, which would put an end to the injustice that social inequality so often generates, succeeds only in destroying the interest of life by denying or attempting to suppress our inevitable human diversity. An authentic and “natural” aspiration to some good expresses itself in the form of an authoritative direction of the more passive members of a society and shapes their lives in the light of this aspiration’s own moral vision. Accordingly, Santayana frequently tended to identify strong authoritarian government with the natural bent of a self-assertive vitality and uniformly treated liberalism as an incoherent and sterile principle of dissolution, roughly comparable in its inspiration and effect to the Protestant principle in the province of religion. Both liberalism and the Protestant principle are expressions of that romantic individualism that Santayana was willing to tolerate as a kind of playful self-deception of the “inner life,” but which he abominated whenever it took itself seriously and became a principle of action directed toward correcting the “natural” order of things.

**MORALITY.** Strangely enough, it is in *Reason in Science*, Volume V of *The Life of Reason*, that Santayana’s fullest exposition of his views on morality is to be found. In this work he distinguished between “rational” morality and the morality that is either “prerational” or “postrational.” Rational morality is no longer the straightforward hedonism of *The Sense of Beauty*, for Santayana now recognized that there must be a principle of selective preference among possible enjoyments. But he still regarded our adoption of such an ideal standard as a matter of temperament and natural inclination; and even the attempt to achieve a comprehensive integration of diverse satisfactions, which is what distinguishes rational morality, is presented as just one possible attitude toward life. Rational morality and the moral philosophy associated with it, Santayana argued, are concerned with what is really good, and they require a highly developed capacity for sympathetic understanding and assessment of all

competing goods; but in the end, what is really good can only be what genuinely expresses some vital bias of our natures. By contrast, prerational morality is the unreflective life of primary impulse, which cannot conceive the possibility of alternative goods nor support the discipline entailed by a principled organization of the moral life. Postrational morality, finally, is an essentially religious abandonment of the hope for a rational ordering of human life in favor of some otherworldly ideal. Its sole strength, as Santayana observed, lies in the remnant of natural assertiveness that survives in its condemnation of the works of the natural man and the desperate energy with which a single and exclusive regimen of life is proclaimed to be the sole means of salvation.

**SCIENCE.** Santayana's attitude toward science, as one phase of the life of reason, was an inconsistent mixture of hospitality and indifference. Convinced as he was that all causal efficacy belongs to physical nature, he was strongly inclined to accept the claim of science to exclusive authority in the determination of what is really true. Natural science is at once an extension of common sense and a uniquely successful application of "dialectics," that is, the logical elaboration of terms of thought, or "concretions in discourse," to the study of the physical world. The ideal of such a science would be a closed, mechanistic, and materialistic system, and Santayana believed that progress in the sciences of man, notably psychology, required the adoption of this ideal. But beyond this recognition of the authority of science, Santayana had no detailed interest in its findings and only a very limited belief in its power to contribute to those ideal values that are the true substance of the life of reason. It deals, after all, with only one of many possible worlds; and while the discipline of fact to which it subjects the mind is infinitely preferable to the projection upon the world of some moral fable of our own devising, the highest form of intellectual freedom is still to survey the field of ideal possibilities without any sense of an obligation to describe or a fear of misdescribing any actual state of affairs.

#### SCEPTICISM AND ANIMAL FAITH

In *Scepticism and Animal Faith* (1923), Santayana undertook the extensive recasting of his whole system of thought; to which reference has been made above. The reformulation was to consist in the substitution of a set of ontological distinctions for the introspective psychology of his earlier writings. Properly speaking, this work is an introduction to, and a partial summary of, the main doctrines of *Realms of Being* (1927–1940). It begins with an attempt to radicalize, and thus to overcome, the idealistic

skepticism concerning the existence of an external world that has been a central theme of Western philosophy since René Descartes. The argument is that if we limit ourselves to what is immediately given (and therefore incapable of being doubted), not only our belief in an external world, but also our belief in the existence of the self, of other selves, and of a past and a future is undermined. All that remain are certain characters or essences that bear no relationship to things or events and cannot properly be said to "exist."

Santayana's point is that a genuine skepticism, pushed to its logical extreme, is just as fatal to the "mind" of the idealists as it is to the matter they were prepared to abandon. In a positive sense, the upshot of such skepticism is to reveal essence as the primary and incontestable mode of being; but it is practically and psychologically impossible for human beings to recognize only essential being. "Animal faith" thus supervenes upon the intuition of essence and posits the existence of a world of things and events that transcends immediate intuition. In one sense this belief is quite baseless, since there cannot, in a strict sense, be proof that anything exists; but in another sense this belief is the beginning of wisdom. In this conception there is no great shift away from the view set forth in *The Life of Reason*. The chief difference, however, is that in *Scepticism and Animal Faith* the commitment to existence and to substance (which in the earlier work was presented retrospectively as the first great achievement in the history of consciousness) is first dramatically revoked and then reinstated by the individual mind. But with respect to the logical status and practical necessity of this belief, Santayana's views would not appear to have undergone any significant change.

#### THE REALMS OF BEING

*The Realms of Being* is a detailed characterization of the four major modes of being or basic categories that emerge from the skeptical self-interrogation of consciousness. The modes of being consist of essence and matter, as noted above, and two derivative modes, truth and spirit.

**ESSENCE.** The being of essence is first carefully distinguished from certain adventitious notions that have been associated with it in the history of Western philosophy. Among these are the views that attribute causal efficacy or some special moral or aesthetic status to essences as such, and also the views that envisage essence only in the context of some mental activity such as "abstraction" or "imagination." Properly conceived, the being of essence

consists simply in the self-identity of its character. Since this intrinsic character involves no reference to any location in space or time, essences are universal and repeatable. They are infinite in number and yet collectively compose one absolute essence in “Pure Being,” which is common to all essences. Essences are logically discrete and individual, and one essence can “imply” another only if it is first stipulated that the relationship is that of a whole to one of its parts and that no logical necessity governs the constitution of such wholes. Essences may be exemplified in the realm of matter, but they need not be; and even when they are, the things and events that are the bearers of these ideal characters have a quite different mode of being.

**MATTER.** The “indispensable properties” of the material mode of being are spatial extension and temporal process. Matter exists contingently and is therefore unstable and evanescent; but it also maintains a dynamic continuity, through change and can in this sense be called “substance.” It is external to and independent of consciousness; and it is ultimately unknowable, since we know it only through the essences it exemplifies—and these are radically incapable of representing the element of process and diffusion that is peculiar to the realm of matter. Organisms are part of that realm, and the psychological histories (as distinct from the pure consciousness) of human beings can be understood only by reference to the behavioral unity that Santayana calls the “psyche.”

**TRUTH.** Originally, Santayana had intended to establish only three “Realms of Being,” and in fact the Realm of Truth that he later added has obvious affinities with both essence and matter. Truth is the truth about matter, or what exists, and yet it is independent of existence both because “no fact can be a description of itself” and because even if nothing existed, it would still be true that nothing did exist or that just such and such things had existed in the past. Truth is “the sum of all the propositions,” and as such it represents a certain selection from the infinite essences or character that things might have had. Truth is timeless and independent of all beliefs. There are no necessary truths, and even the propositions of mathematics are only contingently true since it is simply an accident if they correctly describe the material world.

**SPIRIT.** Spirit, as Santayana used the term, is simply pure transcendental consciousness, and as such it must be distinguished from its physical basis (the “psyche”) and from particular mental events. The only criterion of the exist-

tence of spirit is internal; and it exists contingently. It is entirely passive in its relation to physical nature, and its sole function is pure intuition, which, Santayana says, is “the direct and obvious possession of the apparent without commitments of any sort about its truth, significance, or material existence.” The unities of intuition are simply individual essences and are not the product of any mental machinery. By itself, intuition is not cognitive. Considered simply as a skein of meanings, the life of intuition may acquire a unity and a life and even a kind of freedom that lacks the power to intervene in the world but is nevertheless the highest and purest human good.

To some extent, *The Realms of Being* effects a clarification of Santayana’s earlier views, although it may be doubted whether he was ever in much danger of being taken for an idealist. Unfortunately, the style of the later book is even more luxuriant than that of *The Life of Reason*, and Santayana’s unwillingness to argue technical philosophical issues was still as strong as ever. If what he hoped to present in *Realms of Being* was, as he says, a language in which the great distinctions to which we all have recourse would be clearly marked out, his success must be judged to be only very partial. All doctrines of transcendental subjectivity, including Santayana’s, engender immense difficulties which cannot be resolved unless the philosopher is more inclined to meet criticism on some ground other than the assumed truth of his own views. In *Realms of Being*, there are very few signs, of such a disposition on Santayana’s part.

## CRITICAL WORKS

Santayana was not just a philosopher in his own right but also a critic, both philosophical and aesthetic. Several of his books, among them *Interpretations of Poetry and Religion* (1900), *Three Philosophical Poets* (1910), *Winds of Doctrine* (1913), *Character and Opinion in the United States* (1920), *Platonism and the Spiritual Life* (1927), and *Obiter Scripta* (1936), are made up of critical studies of systems of thought as diverse as the pragmatism of William James and the atomism of Lucretius; and in many ways, Santayana was at his best as a critic; and in many ways, Santayana was at his best as a critic. In spite of the severity of his judgments and his tendency to use both philosophers and imaginative writers as stalking horses for his own philosophical purposes, he seldom failed to make some telling observation or incisive criticism that had a validity independent of his own special point of view. At the same time, it must be noted that in his critical essays he too often affected an Olympian manner that only partially concealed the strongly personal

character of his tastes and distastes both for individuals and ideas.

**See also** Aesthetic Judgment; Aesthetics, History of; Art, Expression in; Beauty; Descartes, René; Essence and Existence; Husserl, Edmund; Kant, Immanuel; Lotze, Rudolf Hermann; Realism; Russell, Bertrand Arthur William; Skepticism, History of; Spinoza, Benedict (Baruch) de; Value and Valuation.

## Bibliography

### WORKS BY SANTAYANA

#### Major Philosophical Works

*The Sense of Beauty*. New York: Scribners, 1896.

*The Life of Reason; or The Phases of Human Progress*. 5 vols.

New York: Scribners, 1905–1906; 1 vol. ed., New York, 1954.

*Scepticism and Animal Faith*. New York: Scribners, 1923.

*Realms of Being*. 4 vols. New York: Scribners, 1927–1940; 1 vol. ed., 1941.

“Three Proofs of Realism.” In *Essays in Critical Realism*. New York, 1941.

*Dominations and Powers*. New York: Scribners, 1949.

#### Critical Writings

*Interpretations of Poetry and Religion*. New York: Scribners, 1900.

*Three Philosophical Poets*. Cambridge, MA: Harvard University Press, 1910.

*Winds of Doctrine*. New York: Scribners, 1913.

*Egotism in German Philosophy*. New York: Scribners, 1915.

*Character and Opinion in the United States*. New York: Scribners, 1920.

*Platonism and the Spiritual Life*. New York: Scribners, 1927.

*Some Turns of Thought in Modern Philosophy*. New York: Scribners, 1933.

*Obiter Scripta*, edited by J. Buchler and B. Schwartz. New York: Scribners, 1936.

*The Idea of Christ in the Gospels*. New York: Scribners, 1946.

#### Other Works

*The Last Puritan*. New York: Scribners, 1936. Santayana’s only novel, which is also a “memoir” of considerable interest.

*Persons and Places*. 3 vols. New York: Scribners, 1944–1953. Autobiography.

*The Letters of George Santayana*, edited by Daniel Cory. New York: Scribners, 1955.

### WORKS ON SANTAYANA

Duron, Jacques. *La pensée de George Santayana*. Paris: Nizet, 1950. A massive and detailed study of Santayana’s whole philosophy.

Howgate, George W. *George Santayana*. Philadelphia: University of Pennsylvania Press, 1938. A less technical study of Santayana, mainly as a critic and aesthetician.

Munitz, Milton. *The Moral Philosophy of Santayana*. New York: Columbia University Press, 1939.

Schilpp, P. A., ed. *The Philosophy of George Santayana*. Vol. II in the Library of Living Philosophers. Evanston, IL: Northwestern University Press, 1940. A collection of critical

essays by a group of philosophers, among them Bertrand Russell, with an autobiographical statement by Santayana, a detailed reply to his critics, and a detailed bibliography. The most useful single work on Santayana’s philosophy.

Singer, Irving. *Santayana’s Aesthetics*. Cambridge, MA: Harvard University Press, 1957.

*Frederick A. Olafson (1967)*

## SARTRE, JEAN-PAUL

(1905–1980)

Jean-Paul Sartre, French existentialist philosopher and author, was born in Paris where he attended prestigious *lycées* and then the *École Normale Supérieure* from 1924 to 1928. After passing his *agrégation* the following year, he taught in several *lycées* both in Paris and elsewhere. In 1933, he succeeded Raymond Aron (1905–1983) as a research stipendiary for a year at the Institut Français in Berlin, where he immersed himself in phenomenology, concentrating on Edmund Husserl but also reading Max Scheler and some Martin Heidegger. In the years following his return to France, he published several phenomenological works as well as the philosophical novel *La nausea* (*Nausea*) (1938) that brought him public recognition. He resumed his teaching till conscripted into the French Army in 1939. After serving ten months as a prisoner of war chiefly in Trier, where he taught Heidegger’s *Being and Time* (1962) to several imprisoned priests and continued writing his masterwork caps for *L’être* (*L’être et le néant*) (Being and nothingness) (1943), he returned to Paris for three more years of *lycée* teaching. Soon he was able to make his living from his writing and would never teach again. He was involved in a short-lived resistance movement of intellectuals that included Maurice Merleau-Ponty and Simone de Beauvoir, the latter his lifelong companion. With these and several others, he founded the journal *Les temps modernes* (Modern times), its first issue appearing in October 1945, which quickly became the voice of existentialism and remains a leading literary and political publication to this day.

In the aftermath of the war, Sartre emerged as the leader of the existentialist movement, the quasi manifesto of which he delivered in a famous address subsequently published as *L’existentialisme est un humanisme* (1946). By then, he was world famous. He used his celebrity to promote political and social causes of the Left in accord with the theory of *committed literature* introduced in a series of essays published as *Qu’est-ce que la littérature?* (What is literature?) in *Les temps modernes* (1947). He wrote a number of short stories, novels, and plays as well

as several studies of the lives of famous authors, including his autobiography, *Les mots* (The words), for which he was awarded the Nobel Prize in Literature (1964) and which he declined. After an unsuccessful association with an incipient noncommunist nonparty of the Left, he abandoned organized politics. His relations with the Communist Party ran hot and cold. Initially vilified by the party as a bourgeois individualist, he gradually became a fellow traveler, using different standards with which to judge the East and the West during the Cold War. But after the Soviet occupation of Budapest in 1956 and the crushing of the Prague Spring in 1968, he turned against the French Communist Party and moved farther Left, titling one interview: “*Les Communistes ont peur de la révolution*” (The Communists are afraid of revolution) (*Situations*, VIII, 1969). In 1960 he published his second major philosophical work, the first volume of *Critique de la raison dialectique, précédé de questions de méthode* (The critique of dialectical reason) preceded by a kind of preface *Questions de méthode* (Search for a method) that had appeared in *Les temps modernes* in 1957. This marked his theoretical shift from a philosophy of consciousness and subjectivity to one of dialectical praxis (human activity in its socioeconomic milieu). Many see this as the theoretical basis for the student revolt known as *the events of May, 1968* that constituted a turning point in French cultural life.

Throughout these years of political turmoil and despite his proclaimed abandonment of imaginative literature in favor of political action, Sartre continued to labor on his multivolume study of Gustave Flaubert’s life and times, *L’idiot de la famille; Gustave Flaubert de 1821–1857* (The family idiot: Gustave Flaubert de, 1821–1857) (1971–1972). After a number of strokes in the 1970s left him almost totally blind, he began a series of interviews with former Maoist activist Benny Lévy (1945–2003), then serving as his secretary, that he announced would leave none of his earlier positions unchanged. The proposed elements of an *ethic of the ‘We’*, as he called it, appeared in three issues of the weekly magazine *Le nouvel observateur*. Titled “*L’espoir maintenant*” (Hope now) these interviews constitute his last publication during his lifetime. After his death on April 15, 1980, the funeral cortege was joined by thousands of people in the largest spontaneous demonstration Paris had seen since the death of France’s president Charles De Gaulle (1890–1970). France had lost “the conscience of his time,” proclaimed the lead essay in a major journal (*Magazine littéraire*, September 1981) and the immense crowd of mourners seemed to agree.

## A PHILOSOPHER OF THE IMAGINATION

Starting with his thesis for the diplôme d’études supérieures titled “The Image in Psychological Life: Role and Nature” (1926) Sartre exhibited a strong interest in the realm of the imaginary. This becomes the object of two of his early publications, *L’imagination* (1936), a reworking of the earlier thesis, and the more important *L’imaginaire* (The imaginary) (1940), in many ways the key to his subsequent thought. For what he attributes to imaging consciousness in the latter—namely, that it is the locus of possibility, negativity, and lack—is precisely how he will later characterize being-for-itself or consciousness in *Being and Nothingness*. Imaging consciousness becomes the paradigm of consciousness in general for Sartre.

From this follow several characteristic features of his aesthetics, ethics, and political theory as well as the *choice* of the imaginary on the part of the subjects of his existentialist biographies or *psychoanalyses*. It also explains the ease with which he employed the method of *free imaginative variation of examples* (eidetic reduction) from Husserlian phenomenology in constructing his philosophical position. Many of his *arguments* are descriptive in nature, exhibiting Husserl’s remark that the point of phenomenology is not to explain but to get us to *see*. Moreover, the matching of imaging consciousness with conceptual analysis in Sartre’s works serves to bridge the commonly perceived distance between philosophy and imaginative literature, helping us better appreciate the philosophical approach to literature and the literary approach to philosophy that mark his writings. His novel *Nausea*, for example, anticipates, and his play *No Exit* (1944) applies, theses and themes of *Being and Nothingness* in concrete fashion.

Sartre remained faithful to the descriptive method of phenomenology throughout his career. Even when he introduced the dialectical *progressive-regressive* method in *Search for a Method*, it was to be preceded by a phenomenological description of the situation at hand. But he was not an uncritical reader of Husserl. In a major essay, “Transcendence of the Ego,” composed while in Berlin but published in 1937, Sartre defends what Aron Gurwitsch called a *nonegological conception of consciousness*. The *of* in the title denotes both a subjective and an objective genitive: The transcendental ego of Husserlian phenomenology has been rendered unnecessary (transcended) whereas the empirical ego (the subject of our reflective knowledge and scientific study) *transcends* consciousness in the sense that it is *other* than the conscious-



ness one has of it. This allows Sartre to distinguish between an autonomous, *prereflective* consciousness that is impersonal or *prepersonal* and the realm of reflective awareness that constitutes our psychological life, which he will call the *Psyche*. He wrote a lengthy manuscript on the latter, only a portion of which was ever published—*Esquisse d'une théorie des émotions* (Sketch for a phenomenological theory of the emotions) (1939).

One of the core theses of phenomenology is the claim that all consciousness is consciousness of an other-than-consciousness. Consciousness simply is this aiming at or *intending* an object. This is Husserl's famous thesis of *intentionality* as the defining characteristic of the mental. Perhaps no other phenomenologist has pursued the thesis of intentionality with such consistency as Sartre, even to the point of accusing Husserl, rightly or not, of having betrayed this principle by his understanding of mental images as simulacra *inside* the mind. Sartre will insist that if images are conscious, then they, too, are ways of intending the world as are our emotions. The challenge is to articulate the distinguishing features of these various ways of being *in-the-world*, an expression Sartre adopts from Heidegger.

In *The Imaginary* Sartre undertakes the task of describing the defining characteristics of the image. Relying on the evidence from his reflective description of our prereflective awareness, he identifies four essential features of the image:

1. *The image is a consciousness* rather than an object *inside* consciousness (Sartrean consciousness has no inside; it is essentially *outside*, in-the-world). The image is a relationship to an object. Hence, it is more accurate to speak of *imaging consciousness* than of *images*. The latter term suggests miniatures that we project outside the mind, an example of what Sartre terms *the illusion of immanence*, which is contrary to the intentionality of consciousness.

2. In contradistinction to perception, which must grasp its object in *profiles* that it synthesizes into a perceptual judgment of identity (these are profiles of one and the same cube that cannot all be given simultaneously) imaging consciousness presents its object all at once (we see in the object only what we place there; the image teaches us nothing). Whereas the perceived object *overflows* our perception of it and invites further investigation, in the case of imaging consciousness, what you imagine is what you get. The *studying* of an imagined object is actually the sequential *viewing* of a series of imagings. Sartre calls this the phenomenon of *quasi-observation*. I can syn-

thesize the series into the object of flesh and blood (my friend Peter, for example) that I could perceive, were he available for perception, but ex hypothesi, as imagined, he is unavailable.

3. *Imaginative consciousness posits its object as a nothingness*. Sartre describes this as making its object *present-absent*, that is, present but out of the circuit of my perceptual beliefs that define the *real*. The realm of the imaginary is what Sartre designates the *irreal* as distinct from the *unreal*, which could apply to the perceptual or the conceptual realm. Following Husserl, Sartre allows for just four types of presence-absence: One can imagine the object as nonexistent (unicorns), as absent (Peter as not here), as existing elsewhere (Peter in Berlin), or in a neutral mode that simply prescind from its existence (as with ideal objects, for example). This is what distinguishes my awareness of the imagined tree from that of the perceived one, which is grasped as present in its *materiality*. Sartre will elaborate this *nothingness* when he describes the *othering* or *nihilating* nature of consciousness in general in *Being and Nothingness*.

4. Imaging consciousness is *spontaneous*, another feature that Sartre will later extend to consciousness sans phrase. This characteristic denotes the prereflective and implicit (Sartre calls it *nonthetic*) awareness that imaging consciousness has of its creative power as it sustains the object in presence-absence. Sartre will speak of this as an awareness of freedom, which he already extends to prereflective consciousness across-the-board and which he will later liken to Descartes's notion of God's power to conserve in existence the created world.

Much of Sartre's aesthetic theory turns on this idea of the image, which he defines as: "an act that aims in its corporeality at [intends] an absent or nonexistent object, through a physical or psychic content that is given not as itself, but in the capacity of analogical *representative* of the object aimed at [intended]" (Sartre 1940/2003, p. 22). As intentional, consciousness has no contents but it does have objects. In the case of aesthetic objects such as the portrait of Charles VIII or the playing of the Appassionata Sonata, the artifact, say the physical painting or the musical performance, serves as *analogon* for the creative imagination of artist and public alike. By our assuming the *aesthetic* attitude, that is, by *derealizing* the perceptual object, the artifact serves as analogon for making present-absent (re-presenting) this particular aesthetic object. Sartre emphasizes that the imaging act is a synthesis of cognitive and emotional intendings. But his analysis

attends chiefly to the primary role of imaging consciousness in this *derealizing* act.

To indicate the pervasiveness of imagination in Sartre's thought, it suffices at this point to mention the role reserved in his existentialist ethic for the *image* of the kind of person I want to be that is implicit in my moral choices, a clear reference to the phenomenological ethics of Scheler. Nor should we overlook the guiding ideal of the *City of ends* throughout Sartre's political philosophy. And when we recall its character as the locus of negativity, possibility, and lack, the presence of the imagination appears as far-ranging as consciousness itself.

## AN ONTOLOGIST

Sartre remarked late in his career that what distinguished him from the Marxists was that he raised the class question starting with being, which is wider than class, whereas they do not. He elaborates his ontology in two major works.

**BEING AND NOTHINGNESS.** The subtitle of *Being and Nothingness* is "A Phenomenological Ontology." Like Heidegger, whose presence is palpable in this work as was that of Husserl in the earlier ones, Sartre begins his study with the being for whom being is a problem, namely, *human reality* (Heidegger's *Dasein*). By accepting this translation of that basic Heideggerian term, Sartre already seems to be following the anthropological track that Heidegger sought to move beyond. But, in fact, Sartre, too, is concerned with gaining access to being in order to delineate its fundamental modes. Still, his point of access is the immediate experience of the *phenomenon of being* in experiences of boredom, nausea, and the like.

In his novel *Nausea*, Sartre's protagonist experiences the sheer contingency of the tree root that captures his attention, its gratuitous *existence*—and his own: "Every existing thing is born without reason, prolongs itself out of weakness and dies by chance" (*Nausea* 1964, p. 113). Sartre's formal ontology in *Being and Nothingness* will follow from the descriptive analysis of that phenomenon of the being of things. Against idealism, against those who succumb to the *illusion of immanence*, Sartre insists on the *transphenomenal* character of being, that is, its irreducibility to appearances. Showing himself as much the pupil of Henri Bergson as of Heidegger in this regard, Sartre appeals to a *revealing intuition of the phenomenon of being*. But this being is not some Kantian thing-in-itself standing behind the appearances; the phenomenon of being is coterminous with, though irreducible to, the being of the phenomena. The phenomena that the eidetic

reduction yields are the objects of knowledge; for example, the kind of knowledge that we gain about the nature of imaging consciousness. Such phenomena are reflective and our awareness of them cognitive. The phenomenon of being is prereflective and noncognitive. It follows that knowledge cannot give an account of transphenomenal being. To attempt to do so Sartre calls *metaphysics*, to which he gives short shrift toward the end of the book.

Using the phenomenological method of descriptive analysis, Sartre discovers three irreducible modes of being, namely, being in-itself, or the inert; being-for-itself, or the spontaneous (consciousness); and being-for-others, or the interpersonal. Though he claims that the for-others is as fundamental as the for-itself, it is clear that being-for-others is inconceivable without the other two, which are conceivable without it. So having distinguished between being and the phenomena, Sartre's descriptive analysis now reveals two radically different regions of being: the transphenomenal *being* of the prereflective cogito or *I think* that precedes and sustains any reflective awareness such as Descartes's *Cogito* or any other phenomena insofar as they are consciousness-relative, on the one hand, and the transphenomenal being of the objects of consciousness, revealed in the experiences of nausea, boredom and the like, on the other.

Pursuing this analysis, Sartre discovers that consciousness, which he will soon call *being-for-itself*, simply is the transphenomenal dimension of nonbeing, which he calls *nothingness* (*le néant*), the nothingness of *Being and Nothingness*, whereas *being-in-itself* denotes the dimension of transphenomenal being of the object of consciousness. Each region bears distinctive features. Being in-itself, in Sartre's metaphorical discourse, is thing-like in its solidity and identity. An inert plenum, the in-itself simply is what it is. This region includes any aspect of experience that manifests these properties; for example, substances or the temporal past or any of the *givens* of our experience that Sartre, borrowing from Heidegger, calls our *facticity*. Once other subjects enter the scene and a third, irreducible, dimension emerges, which Sartre calls *being-for-others* (*l'être-pour-autrui*), the scope of facticity expands to include such givens as our reputations, social institutions, and cultural phenomena generally. These, too, are forms of being-in-itself.

Being-for-itself bears contradictory features. As the nothingness of *Being and Nothingness*, the for-itself is the internal negation, or *nihilation*, of the in-itself. Sartre agrees with Heidegger that negativity is not simply a property of propositions but that it is introduced into the world by human reality itself. As evidence, Sartre cites a

whole series of *negativities* (*négativités*), such as our experience of the *fragility* of entities, of absence, of distance, of distraction, of regret, and of lack. (Recall his characterization of imaging consciousness). The for-itself is an exception to the Parmenidean rule of self-identity: Consciousness is nonself-identical. It is always *other* than itself, which is an ontological expression of its intentionality.

That *inner distance* that separates consciousness from itself accounts for three major characteristics of *human reality* (which is the human being as a composite but not a synthesis of these two ontological regions, related as thing and no-thing). First, it gives rise to the three dimensions of original, *ekstatic* temporality whereby human reality *stands out* from the other and from its very self, namely, the past as facticity, the future as existence or project, and the present as presence-to. This is another way of parsing the nonself-identity of the for-itself. A second consequence of this *gap* or *time lag* that consciousness introduces is the ontological freedom that characterizes our existence. Human reality is free, Sartre insists, because it is not a self but a presence-to-self. Part of Sartre's political endeavor after the war is to pursue the kind of *concrete freedom* that completes this abstract freedom as the definition of the human. Finally, it is this nonself-coincidence that accounts for the paradoxical discourse that Sartre adopts with regard to human reality. Besides the traditional paradoxes of temporality that he inherits, the chief paradox is that human reality *is what it is not* (its possibilities) and *is not what it is* (its facticity as *nihilated* by consciousness). On this account, whatever I am, be it my previous choices or the labels others have affixed to me, I am in the manner of *not-being* them, that is, with the possibility of changing my particular stance in their regard. For the quasi motto of Sartrean humanism is that *you can always make something out of what you've been made into*. This is both the burden of our responsibility and the source of our hope.

With the advent of another subject into my world comes another realm of being as well—being-for-others. Ontologically, this gives rise to an additional set of characteristics that belong to the interpersonal dimension of our existence. The existence of the other subject cannot be deduced; it must be encountered. The most dramatic argument for the existence of other subjects is Sartre's eidetic reduction of shame consciousness. His descriptive analysis centers on the experience one has of being caught in the act of looking at a couple through a keyhole. The feeling of shame that registers in bodily changes such as the face turning red is stronger evidence for the existence

of other minds, Sartre believes, than any argument from analogy. As he unpacks the experience, in one and the same moment, I become aware of the vulnerability of my embodiedness to the look of the other. In other words, what is revealed in this instant is my prereflective consciousness of being objectified by that gaze of another subject. My experience of objectification is simultaneously my experience of the other as subject. Even if on this occasion I happen to be mistaken about the source of the sound I hear behind me, the experience is indicative of being seen by another.

Though Sartre admits that other, derivative modes of access to being-for-others are available (for example, the existence of cultural objects such as directional signs or language itself), he insists that *the look* (*le regard*) is the basic form of interpersonal relation, and he interprets this gaze as objectifying and alienating. "Conflict is the original meaning of being-for-others" (Sartre 1943/1956, p. 364). The interpersonal is like a game of mutual stare-down, each trying to objectify the other. The only type of social philosophy that one can expect from such a thesis is a Hobbesian *war of all against all*. In a famous footnote Sartre concedes that "an ethic of deliverance and salvation" is possible but that this can be achieved only after "a radical conversion" which, he insists, cannot be discussed in that work (Sartre 1943/1956, p. 412). In fact, the elements of an ethic of authenticity are sketched in his posthumously published *Cahiers pour une morale* (Notebooks for an Ethics) composed in 1947–1948, where the basics of this *conversion* are discussed.

Human reality is being-in-situation. *Situation* is composed of facticity and freedom as transcendence; that is, the given that we are always surpassing in our projects. Though he insists that the situation is an ambiguous phenomenon because the precise contribution of each component cannot be determined, it is clear that, as Sartre's sense of social conditioning increases with his shift from abstract to concrete freedom, his respect for the force of circumstance in our situations grows apace. At this stage of his thought, he seems ambivalent as to the limiting and conditioning role of facticity in our actions. But later in life Sartre's sense of what Max Weber called *objective possibility* will heighten and, with it, the claim that fundamental changes in our socioeconomic system are required for abstract freedom to be made concrete. Thus, he will note shortly after the end of the war that "it is the elucidation of the new ideas of 'situation' and of 'being-in-the-world' that revolutionary behavior specifically calls for" ("Materialism and Revolution," Michelson 1962, p. 253).

It is in the context of situation that the concept of bad faith arises. Bad faith is a kind of self-deception, a sort of lying to oneself about the truth of one's situated being. Its most common form consists in collapsing our transcendence (our freedom) into our facticity by appeal to a type of determinism or by simply confessing: *That's just the way I am*. It is a denial of the possibility that consciousness brings to every situation. A related version of this type appeals to the image I wish to present to others or the one they have of me. That, too, is part of my facticity with which I seek to identify in self-deception as if my consciousness did not resist any attempt at full identity. A less common form of bad faith *volatilizes* our facticity into transcendence by choosing to ignore the givens of our situation. This is the bad faith of the dreamer or of the person who flees their past as if it were not part of their situation. But the possibility for self-deception arises from the *dividedness* of our consciousness as prereflective and reflective such that one can be prereflectively aware of more than one knows at the reflective level. Not that one is dealing with two consciousnesses: This deception occurs within the unity of one and the same consciousness.

Since Sartre denies the existence of the Freudian unconscious as he understands it because of its incompatibility with the ontological freedom of human reality, this notion of bad faith cannot appeal to unconscious drives or complexes. What Sartre calls *existential psychoanalysis* aims at dealing with such phenomena as bad faith and fundamental project without appealing to unconscious motives. Its basic premise is that "*man is a totality and not a collection*" (Sartre 1943/1956, p. 568). In other words, at the base of human reality is a fundamental, unifying *choice* that establishes the criteria for all subsequent selections.

We come on the scene having already made that choice, which Sartre believes is guided by the ruling value to consciously be self-identical, that is, to be in-itself-for-itself—an ontological impossibility. This is the meaning of Sartre's famous claim that humankind is a *futile passion*. But how each one lives out that self-defining choice is revealed in the subsequent choices that define a life. "There is not a taste, a mannerism, or a human act," Sartre insists, "which is not revealing" (Sartre 1943/1956, p. 568). The task of psychoanalysis is hermeneutical: to interpret the specific nature of that fundamental choice, that is, the way one acquiesces in or resists that futile passion, by deciphering the symbols of a person's life. What he calls the possibility of *conversion* is the constant threat of altering this basic choice, which haunts our lives. Echo-

ing Danish philosopher Søren Kierkegaard, Sartre calls this the anguish that accompanies the experience of our radical freedom. Admitting that this psychoanalysis has yet to find its Freud, and with a nod toward the work that will occupy a good part of his remaining years, Sartre finds the intimations of such psychoanalysis in certain *successful biographies*.

In many ways, one can read *Being and Nothingness* as an argument moving from the highly abstract (nihilating consciousness, being in-itself and for-itself) to increasingly concrete phenomena such as my *concrete relations with others*, and culminating in the hermeneutic of our particular actions in order to determine the fundamental choice that defines the unity of our lives. Existentialist psychoanalysis both brings this undertaking to a close and opens the door for its application in the several *biographies* that will occupy Sartre's attention over the following decades.

**THE CRITIQUE OF DIALECTICAL REASON.** It was during the war, Sartre insists, that he discovered the philosophical significance of social relations. *Being and Nothingness*, with its emphasis on the *looking/looked-at* model of interpersonal relations, was incapable of explaining the positive reciprocity, collective action, and unintended consequences that a social philosophy requires. In fact, *Being and Nothingness* describes the *we subject* as a "purely subjective *Erlebnis* (experience)" (Sartre 1943/1956, p. 420). Sartre breaks the barrier that confined *Being and Nothingness* to the psychological by introducing the concepts of dialectical praxis, the practice-inert and the mediating third. Together, they account for the dialectical enrichment of individual praxis by group praxis that bears properly social predicates such as rights/duties, power, and function while preserving the freedom and responsibility of the individual, which is a defining characteristic of existentialist thought.

*Praxis* supplants *consciousness* in the lexicon of the *Critique*. It denotes human activity in its sociohistorical context. Praxis is dialectical in the sense that it both negates and conserves aspects of its object in a totalizing action that advances toward a more comprehensive viewpoint. Thus, the negative reciprocity of two boxers in a match, in Sartre's example, when viewed dialectically, is realizing an enveloping social whole called *professional boxing*, which itself invites a still broader contextualization in various socioeconomic systems, such as racism, colonialism, and capitalism. In Sartre's view dialectical thinking is holistic; unlike analytical reason, it welcomes properly social phenomena as irreducible to purely atom-

istic, usually psychological, relations. While admitting the validity of analytical reason within its domain, at a certain level of abstraction, he notes, the class struggle can be seen as the conflict of rationalities.

Sartre reserves a threefold primacy for free organic praxis in his social ontology: ontological, epistemic, and moral. Ontologically, there are only individuals and real relations among them. Praxis is the constitutive dialectic of social phenomena, which are relational entities constituted by individual praxes. This is true even of group praxis, which is the synthetic enrichment of individual praxes in relation, mediated by each member as *third* to every other. Epistemically, the intelligibility of the group and of other social units is a function of the intelligibility of individual praxis, which is its foundation. Sociohistorical intelligibility is dialectical, and the dialectic is grounded in individual praxis. In other words, Sartre denies the existence of a collective consciousness or subject except insofar as it can be seen as a quality of individuals-in-relation.

Sartre speaks of *comprehension* as the *translucidity of individual praxis*. It assumes the clarity that Sartre has reserved for the prereflective cogito in *Being and Nothingness*. The moral primacy of individual praxis follows from the other two forms. Sartre is intent on preserving the moral responsibility of the group members as well as of those he describes as *serialized* by the mediation of worked matter, such as the television-viewing audience or the crowd waiting for a bus. In either case, whether *the same* in group activity and concern or *other* through the separation effected by the mediation of material things, individuals retain moral and not just causal responsibility for the praxis that sustains such relations.

The second basic component of Sartre's social ontology is what he calls the practico-inert. This complex term introduces aspects of being-in-itself into the realm of action. Sartre describes it as "simply the activity of others in so far as it is sustained and diverted by inorganic inertia" (Sartre 1960/1985, p. 556). Not raw nature, but the *practico-inert* is this mediating factor. It includes the sedimentation of prior praxes whether in the form of socioeconomic systems such as colonialism and capitalism or as alienating forms of thought and behavior such as racism, which Sartre calls a *serial idea*. It constitutes the material memory of a society.

Sartre allows for two fundamental kinds of social reality: the active group constituting the *common field* and the effectively separated though ostensibly united (*serialized*) individuals forming the *practico-inert field*. The practico-inert constitutes *fundamental sociality*.

Since Sartre conceives the group as arising through an essential negation of practico-inert seriality, he characterizes the practico-inert ensemble as *the matrix of groups and their grave*. This rich concept is amenable to analytic reason since it is atomistic in nature. But insofar as it occasions counterfinality in the sense that it sustains the boxer's feints and jabs, the conspirator's traps, and the unintended consequences of historical projects, its very antidialectic plays a role in dialectical rationality, conveying the experience of what Sartre calls *dialectical necessity*. Perhaps Sartre's best example of such counterfinality is the flooding and resultant soil erosion caused by Chinese peasants' deforestation undertaken to conserve their land.

But the concept of the mediating third is the key that opens the door to properly group praxis in Sartre's social ontology. There was a concept of the third in *Being and Nothingness*, but this third exercised an objectifying and an alienating function in accord with the looking/looked at model. That concept continues in the *Critique*, where it generates the alienating relations of serial individuals and collectives. But the *mediating* third is a functional concept denoting the group member who is *the same* as the others in common interest and action. As such, it does not objectify or diminish but enriches the responsibility of each in a common practice. Sartre refers to this ternary relation as a *free, interindividual reality*. Simply put, where the practico-inert mediates, human relations are serial; where praxis mediates, these relations are free. And where the practico-inert is modified by material scarcity, Sartre argues, this mediation becomes violent. Such is his bridge between social ontology and history as we know it.

## AN EXISTENTIALIST BIOGRAPHER AND HISTORIAN

In *Search for a Method*, reprinted as a kind of preface to the *Critique* but more properly its sequel, Sartre introduces the progressive-regressive method for investigating social phenomena. This hybrid of existentialist psychoanalysis and historical materialism serves as the model for his later biographies, especially his multivolume study of Gustave Flaubert's life and times, *The Family Idiot*. Sartre studies the socioeconomic and cultural structures of Flaubert's life, particularly as these conditioned the choices available to a would-be literary artist in the second and third quarters of the nineteenth century (the *regressive* movement), the better to chart the spiral of Flaubert's *personalization* as artist, novelist, and finally author of *Madame Bovary* (1956) (the *progressive* stage). The approach is dialectical in its emphasis on the factors that *mediate* these abstract conditions toward their con-

cretization in Flaubert's *choice* of the imaginary, that is, of an artist's life. Indicative of Sartre's increasingly nuanced opposition to the Freudian unconscious is his remark that "everything took place in childhood ... a childhood we never wholly surpass" (Barnes 1968, p. 59–60 and 64).

The dialectic expands to include the *objective spirit* of the age, which Sartre characterizes as *culture as practico-inert*. Using an expression that Aron had employed to describe narrative history in general, Sartre calls *The Family Idiot a novel that is true (un roman vrai)*. Its dialectical interlacing of history and biography render it a properly existentialist approach to historical understanding.

### A MORALIST

If Sartre was a philosopher of the imagination and an ontologist, he was above all a moralist in the French tradition of Duc François de La Rochefoucauld and François-Marie Arouet de Voltaire. His earlier philosophy of consciousness, as well as the primacy of praxis in the social ontology of the *Critique*, are conceived to preserve freedom and responsibility that are the hallmarks of vintage existentialist thought in the midst of impersonal forces, and what Louis Althusser (1918–1990) called *structural causality*. In the hyperbolic mode that he favored, Sartre insisted that *we are without excuse*.

In the course of his life, Sartre developed one ethical theory, sketched a second, and gestured toward a third, in that order. The first and best known is his ethic of authenticity. He describes *authenticity* briefly in *Réflexions sur la Question Juive (Anti-Semite and Jew)* (1946) as "having a true and lucid consciousness of the situation, in assuming the responsibilities and risks that it involves, in accepting it in pride or humiliation, sometimes in horror and hate" (Becker 1995, p. 90). This seems to yield an ethical style rather than a content. It stresses doing rather than being in the sense of embracing my ontological condition, namely, that whatever I am, I am in the manner of not-being it, that is, in nihilating it. I am its *creative unveiling*, with the anguish and joy that accompanies that prereflective awareness.

The ethical content emerges in his novels, stories, plays, and biographies, especially his biography of Jean Genet, *Saint Genet: Comédien et martyr* (Saint Genet, actor and martyr) (1952) and is elaborated in his posthumously published *Notebooks for an Ethics*, which discusses such concepts as good faith, generosity, and positive reciprocity. Maximizing concrete freedom of choice and action becomes an increasingly important moral precept

as Sartre's social sense confronts exploitative systems and oppressive practices after the war.

Exchanging the vocabulary of *Being and Nothingness* for the discourse of the *Critique* in the notes for two sets of lectures and a collection of unpublished reflections from the same period, Sartre sketches a second, *dialectical* ethics that promotes the value of *integral humanity*. This value includes the moral imperative to satisfy human needs by harnessing the practico-inert. Elsewhere, Sartre envisions a socialism of abundance and the new, currently inconceivable, philosophy of freedom that will follow upon it. These lecture notes seem to turn this ideal into an obligation based on the nonnegotiability of basic human needs. In his last discussions with Lévy, he speaks of an ethic of the *we* that will revise many of his previous claims in this regard. However, these recorded remarks were published only in part, and what is available thus far, despite suggestive insights, does not constitute a coherent moral theory. They remain chiefly of biographical interest.

### CONCLUDING OBSERVATIONS

One of the strengths of Sartre's philosophical thought is its insight into the psychological and moral life of individuals and societies. That same gift for imaginative interpretation that fits so well with descriptive phenomenology and makes him a prize-winning novelist and playwright is suspect in the court of conceptual analysis. And once Sartre turns to historical dialectic, the suspicion is compounded. Much of this is simply philosophical bias, which Sartre attempted to address with his distinction between dialectical and analytical reason and their respective logics. But some of it is a reasonable distrust of a lack of rigor evidenced by what Iris Murdoch called Sartre's *great inexact equations*. And then there are his rather extreme political positions and their accompanying moral ascriptions. While one cannot help but admire Sartre's outrage at social injustice and hypocrisy, a remark once reportedly made about Bertrand Russell could be extended to Sartre in this regard: He has the uncanny ability to hit the bull's-eye on the first shot but undermined by a tendency then to splatter all over the target in exaggeration.

Still, Sartre's observations on bad faith and authenticity are now staples in the ethical discourse of our day. And the basic concepts of his social ontology, namely, praxis, the practico-inert, and the mediating third, make a significant contribution that merits the close scrutiny that the prolixity of the *Critique* has denied them. The Cartesian dualism often attributed to Sartre is misap-

plied. His is not a two-substance ontology since only the in-itself is substantial. But a survey of his social ontology in the *Critique* suggests that his dualism is best described as one of spontaneity and inertia, which sends us back to imaging consciousness once more. Perhaps nowhere is the relation between philosophy and imaginative literature more acutely problematized than in Sartre's work. That, too, deserves close attention. Finally, the lessons of Sartrean existentialism speak directly to the renewed interest among our contemporaries in philosophy as a way of life.

**See also** Beauvoir, Simone de; Bergson, Henri; Cartesianism; Descartes, René; Epistemology; Existential Psychoanalysis; Existentialism; Gurwitsch, Aron; Heidegger, Martin; Historical Materialism; Husserl, Edmund; Kant, Immanuel; Kierkegaard, Søren Aabye; La Rochefoucauld, Duc François de; Marxist Philosophy; Merleau-Ponty, Maurice; Murdoch, Iris; Ontology; Parmenides of Elea; Phenomenology; Russell, Bertrand; Scheler, Max; Voltaire, François-Marie Arouet de; Weber, Max.

## Bibliography

For a complete annotated bibliography of Sartre's works, see *The Writings of Jean-Paul Sartre*, edited by M. Contat and M. Rybalka. Evanston, IL: 1973, updated in *Magazine littéraire* 103–104 (1975): 9–49, and by Michel Sicard in *Obliques* 18–19 (1979): 331–347. Rybalka and Contat have compiled an additional bibliography of primary and secondary source published since Sartre's death in *Sartre: Bibliographie, 1980–1992*. Paris: CNRS éditions, 1993.

*Bulletin du Groupe d'Études Sarrtriennes*, the *Yearbook of Sartre Studies*, has published annually since 1987 a list of recently published works of Sartre as well as secondary source material in several languages. Information about this publication, its bibliographies, and the work of the *Groupe* can be found at [www.jpsartre.org](http://www.jpsartre.org).

## WORKS

*L'Imagination*. Paris: Félix Alcan, 1936. Translated by Forrest Williams as *Imagination, A Psychological Critique*. Ann Arbor: University of Michigan Press, 1962.

“La transcendance de l'ego. Esquisse d'une description phénoménologique.” In *Recherches philosophiques* 6 (1936): 65–123. Reprinted in edition by Sylvie le Bon, Paris, 1965. Translated by Forrest Williams and Robert Kirkpatrick as *Transcendence of the Ego: An Existentialist Theory of Consciousness*. New York: Octagon Books, 1972.

*La nausée*. Paris: Gallimard, 1938. Translated by Lloyd Alexander as *Nausea*. New York: New Directions, 1964.

*Esquisse d'une théorie des émotions*. Paris: Hermann, 1939. New edition, 1960. Translated by Bernard Frechtman as *The Emotions. Outline of a Theory*. New York: Philosophical Library, 1948.

*L'imaginaire*. Paris: Gallimard, 1940. Translated by Jonathan Webber as *The Imaginary: A Phenomenological Psychology of the Imagination*. London: Routledge, 2003.

*L'être et le néant*. Paris: Gallimard, 1943. Translated by Hazel E. Barnes as *Being and Nothingness; An Essay on Phenomenological Ontology*. New York: Philosophical Library, 1956.

*Réflexions sur la question Juive*. Paris: P. Morihien, 1946. Translated by George J. Becker as *Anti-Semite and Jew*. New York: Schocken, 1995.

*L'existentialisme est un humanisme*. Paris: Gallimard, 1946. Présentation et notes par Arlette Elkaïm-Sartre. Paris: Gallimard, 1996. Translated by Philip Mairet as *Existentialism and Humanism*. London: Meuthen, 1957. Also translated by Bernard Frechtman as *Existentialism*. Secaucus, NJ: Citadel 1957.

*Situations*. 10 vols. Paris: Gallimard, 1947–1976. Contains reprints of important essays and interviews such as “La Liberté Cartésienne” (Vol. 1, pp. 289–307) and “Matérialisme et Révolution” (Vol. 3, pp. 135–223). Translated by Annette Michelson as “Cartesian Freedom” and “Materialism and Revolution.” In *Literary and Philosophical Essays*. New York: Collier, 1962.

*Qu'est-ce que la littérature? Les temps modernes* 17–22 (1947): no. 17, pp. 769–805; no. 18, pp. 961–88; no. 19, pp. 1194–1218; no. 20, pp. 1410–1429; no. 21, pp. 1607–1641; no. 22, pp. 77–114. Reprinted in *Situations* 2. Paris: Gallimard, 1948. pp. 55–330. Translated by Bernard Frechtman et al. as *What Is Literature? And Other Essays*. Cambridge, MA: Cambridge University Press, 1988.

*Saint Genet: Comédien et martyr*. Paris: Gallimard, 1952. Translated by Bernard Frechtman as *Saint Genet: Actor and Martyr*. New York: G. Braziller, 1963.

*Critique de la raison dialectique, précédé de questions de méthode*. Vol. 1, *Théorie des ensembles pratiques*. Paris: Gallimard, 1960. Reprinted in a new annotated edition, 1985. The prefatory essay to the *Critique*, “Questions de méthode,” appeared in *Les temps modernes* 139–140 (1957): no. 139, pp. 338–417; no. 140, pp. 658–698. Translated by Hazel E. Barnes as *Search for a Method*. New York: Random House, 1968. An annotated edition of *Questions de méthode* by Arlette Elkaïm-Sartre appeared in 1986. Paris: Gallimard, collection Tel. The *Critique* itself was translated by Alan Sheridan-Smith as *Between Existentialism and Marxism (Essays and Interviews, 1959–70)*. Translated by John Mathews. London: NLB, 1974.

*Les mots*. Paris: Gallimard, 1964. Translated by Bernard Frechtman as *The Words*. New York: G. Braziller, 1964.

*L'Idiot de la famille; Gustave Flaubert de 1821–1857*. Vols. 1–3, Paris: Gallimard, 1971–1972. Revised edition of Vol. 3, 1988. Translated by Carol Cosman as *The Family Idiot: Gustave Flaubert, 1821–1857*. 5 vols. Chicago: University of Chicago Press, 1981–1993.

*On a raison de se révolter: Discussions*. With Philippe Gavi and Pierre Victor. Paris: Gallimard, 1974.

*Critique of Dialectical Reason*. Vol. 1, *Theory of Practical Ensembles*. London: NLB, 1976. A new edition based on the revised French edition of 1985 appeared in 2004. London: Verso.

*Sartre on Theater*, edited by Michel Contat and Michel Rybalka, translated by Frank Jelinek. New York: Pantheon, 1976.

*Situations*. Vol. 10. Paris: Gallimard, 1976. Translated by Paul Auster and Lydia Davis as *Life/Situations: Essays Written and Spoken*. New York: Pantheon, 1977.

*Oeuvres romanesques*. Paris: Gallimard, 1981.

*Cahiers pour une morale*. Paris: Gallimard, 1983. Translated by Davie Pellauer as *Notebooks for an Ethics*. Chicago: University of Chicago Press, 1992.

*Les carnets de la drôle de guerre*. Paris: Gallimard, 1983. Translated by Quintin Hoare as *The War Diaries of Jean-Paul Sartre: November 1939–March 1940*. New York: Pantheon, 1984.

*Scénario Freud*. Paris: Gallimard, 1984. Translated by Quintin Hoare as *The Freud Scenario*. Chicago: University of Chicago Press, 1985.

*Critique de la raison dialectique*. Vol. 2, *L'intelligibilité de l'histoire*. Paris: Gallimard, 1985. Translated by Quintin Hoare as *Critique of Dialectical Reason*. Vol. 2, *The Intelligibility of History*. London: Verso, 1991.

*Vérité et existence*, edited by Arlette Elkäim Sartre. Paris: Gallimard, 1989. Translated by Adrian van den Hoven as *Truth and Existence*, edited by Ronald Aronson. Chicago: University of Chicago press, 1992.

*L'espoir maintenant: Les entretiens de 1980*. Lagrasse: Verdier, 1991. Translated by Adrian van den Hoven, as *Hope Now: The 1980 Interviews*. Chicago: University of Chicago Press, 1996.

*Théâtre complet*, edited by Michel Contat. Paris: Gallimard, 2005.

#### SELECTED STUDIES

Anderson, Thomas C. *Sartre's Two Ethics: From Authenticity to Integral Humanity*. Chicago: Open Court, 1993.

Aronson, Ronald. *Sartre's Second Critique*. Chicago: University of Chicago Press, 1987.

Barnes, Hazel E. *Sartre & Flaubert*. Chicago: University of Chicago Press, 1981.

Bell, Linda A. *Sartre's Ethics of Authenticity*. Tuscaloosa: University of Alabama Press, 1989.

Busch, Thomas. *The Power of Consciousness and the Force of Circumstances in Sartre's Philosophy*. Bloomington: Indiana University Press, 1990.

Catalano, Joseph. *A Commentary on Jean-Paul Sartre's Being and Nothingness*. Chicago: University of Chicago Press, 1980.

Catalano, Joseph. *A Commentary on Jean-Paul Sartre's Critique of Dialectical Reason*. Vol. 1, *Theory of Practical Ensembles*. Chicago: University of Chicago Press, 1986.

de Beauvoir, Simone. *Adieux: A Farewell to Sartre*. Translated by P. O'Brian. New York: Pantheon, 1984

de Beauvoir, Simone. *The Force of Circumstances*. Translated by Richard Howard. New York: Putnam's, 1964–1965.

Detmer, David. *Freedom as a Value: A Critique of the Ethical Theory of Jean-Paul Sartre*. La Salle, IL: Open Court, 1988.

Dobson, Andrew. *Jean-Paul Sartre and the Politics of Reason*. Cambridge, U.K.: Cambridge University Press, 1993.

Fell, Joseph P. *Heidegger and Sartre: An Essay on Being and Place*. New York: Columbia University Press, 1979.

Flynn, Thomas R. *Sartre and Marxist Existentialism: The Test Case of Collective Responsibility*. Chicago: University of Chicago Press, 1984.

Flynn, Thomas R. *Sartre, Foucault and Historical Reason*. Vol. 1. *Toward an Existentialist Theory of History*. Chicago: University of Chicago Press, 1997.

Flynn, Thomas R. *Sartre, Foucault and Historical Reason*. Vol. 2. *A Poststructuralist Mapping of History*. Chicago: University of Chicago Press, 2005.

Howells, Christina, ed. *The Cambridge Companion to Sartre*. Cambridge, U.K.: Cambridge University Press, 1992.

Jeanson, Francis. *Sartre and the Problem of Morality*. Translated by Robert Stone. Bloomington: Indiana University Press, 1981.

McBride, William Leon, ed. *Sartre and Existentialism*, 8 Vols. New York: Garland, 1997.

McBride, William Leon. *Sartre's Political Theory*. Bloomington: Indiana University Press, 1991.

Santoni, Ronald E. *Bad Faith, Good Faith and Authenticity in Sartre's Early Philosophy*. Philadelphia: Temple University Press, 1995.

Santoni, Ronald E. *Sartre on Violence: Curiously Ambivalent*. University Park: Pennsylvania State University Press, 2003.

Schilpp, Paul Arthur, ed. *The Philosophy of Jean-Paul Sartre*. La Salle, IL: Open Court, 1981.

Schroeder, William. *Sartre and His Predecessors: The Self and the Other*. Boston: Routledge & Kegan Paul, 1984.

Silverman, Hugh J. *Inscriptions: Between Phenomenology and Structuralism*. London: Routledge & Kegan Paul, 1987.

Stone, Robert, and Elizabeth Bowman. "Dialectical Ethics: A First Look at Sartre's Unpublished 1964 Rome Lecture Notes." *Social Text* 13–14 (1986): 195–215.

Stone, Robert, and Elizabeth Bowman. "Sartre's 'Morality and History': A First Look at the Notes for the Unpublished 1965 Cornell Lectures." In *Sartre Alive*, edited by Ronald Aronson and Adrian van den Hoven, 53–82. Detroit, MI: Wayne State University Press, 1991.

Taylor, Charles. *The Ethics of Authenticity*. Cambridge, MA: Cambridge University Press, 1991.

Verstraeten, Pierre, ed. *Sur les écrits posthumes de Sartre*. Brussels: University of Brussels, 1987.

Thomas R. Flynn (2005)

## SAVAGE, LEONARD

(1917–1971)

Leonard James Savage was the most influential Bayesian statistician of the second half of the twentieth century. Born November 20, 1917, in Detroit, Michigan, Savage received his PhD in mathematics at the University of Michigan in 1941. He then spent a year serving as John von Neumann's assistant at the Institute for Advanced Study in Princeton, where he was exposed to von Neumann's ideas on game theory and the mathematical modeling of human behavior, topics that became a central focus of Savage's research. In his next position at Columbia University's wartime Statistical Research Group—whose members included such luminaries as Abraham



Wald, Milton Friedman, Harold Hotelling, Fredrick Mostler, and Abraham Girshick—Savage developed an interest in statistics and became convinced that the subject should be grounded on a “personalist” conception of probability. After Columbia, Savage went on to hold academic positions at Chicago, Michigan, and Yale.

Savage’s research focused on the mathematical analysis of rational belief and desire, and the advancement of Bayesianism in statistics. His masterpiece, *The Foundations of Statistics* (1954), pursued both these projects by first developing what has come to be the canonical version of subjective expected utility theory, and then attempting to recast all of statistical methodology along subjectivist Bayesian lines.

### SAVAGE’S CONTRIBUTIONS TO DECISION THEORY

Savage’s most notable contributions to the study of rational behavior were his construction of a general framework for modeling decisions under uncertainty, his systematic defense of subjective expected utility maximization as the hallmark of rational choice, and his innovative account of the role of “personal” probabilities in decision making.

Savage portrays decision making as being a matter of using beliefs about possible *states of the world* to choose *actions* that provide the optimal means of producing desirable *consequences*. Actions are identified with functions from states to consequences, and the agent is assumed to have a preference ranking over all acts at her disposal. Influenced by the behaviorism that dominated the social sciences of his day, Savage interpreted preferences operationally, so that an agent may be said to prefer one act *f* to another *g* if and only if she would be disposed to freely choose *f* over *g*. Overt choices thus function as “observables” in decision theory, and talk about the underlying beliefs and desires that cause them is rendered scientifically respectable by showing how they can be operationally defined in terms of preferences. (Savage’s behaviorism remains controversial, but some commentators, e.g., Joyce (1999), regard it as inessential to his overall account of rationality.)

Following Frank Ramsey (1931) and Bruno de Finetti (1937), Savage invoked the hypothesis of subjective expected utility maximization to forge a link between empirically measurable preferences and hidden beliefs and desires. Given a probability function *P* defined over states of the world, and a utility function *u* defined over consequences, the expected utility of an act *f* is the probability-weighted average of the utilities of *f*’s conse-

quences. When there are finitely many states,  $s_1, s_2, \dots, s_n$ , this expected utility is defined as  $Exp_{P,u}(f) = P(s_1)u(f(s_1)) + P(s_2)u(f(s_2)) + \dots + P(s_n)u(f(s_n))$ . Savage maintained that an agent’s preferences can only be deemed rational to the extent that they can be represented as ranking acts according to increasing subjective expected utility.

To establish this conclusion, Savage proposed that any rational preference ranking should satisfy a specific system of axiomatic constraints. The central axiom is the *sure-thing principle*, which states, roughly, that for any acts *f* and *g*, and any event *E*, if *f* is preferred to *g* both conditional on *E* and conditional on not-*E* then *f* is preferred to *g* outright. Savage went on to prove that any preference ranking satisfying his axioms implicitly defines a unique subjective probability *P*, which represents the agent’s degrees of confidence in various states, and a utility *u*, which gauges the strength of her desires for consequences. The agent prefers *f* to *g* just in case  $Exp_{P,u}(f) > Exp_{P,u}(g)$ . In this way, the hypothesis of expected utility maximization allows us to extract degrees of belief and desire from rational preferences.

Many objections to Savage’s theory misinterpret it as a descriptive account, but it was clearly meant to be prescriptive. The most serious doubts about the theory’s normative import concern the status of the sure-thing principle, which some critics see as improperly prohibiting certain sorts of rational aversions to risk or uncertainty. Savage always regarded such worries as misguided, and steadfastly defended the principle’s normative credentials. Many people agree with him, as evidenced by the fact that Savage’s theory, or its close variants, remain central to treatments of rational decision making across the social sciences.

### SAVAGE’S CONTRIBUTIONS TO STATISTICS

Savage maintained that the subjective or “personal” probabilities that figure into decision making should serve also as the basis for statistical reasoning. He implacably opposed the frequentist paradigm that had come to dominate statistics during the 1930s and 1940s. In *Foundations* Savage had tried to incorporate the methods of frequentist statisticians, like Ronald A. Fisher and Jerzy Neyman, into his personalist framework, but by the end of his career he had entirely “lost faith in the devices of the frequentist schools” (Savage 1954). In the second edition of *Foundations* (1972), written six months before his death, he rejects as “ill-founded” such frequentist devices as minimax rules, confidence intervals, tolerance intervals, significance tests, and fiducial probabilities. To take

their place he advocated a thoroughgoing Bayesianism in which all question of statistical reasoning boil down to the choice of a prior personal probability and the use of Bayes's rule to alter personal probabilities in light of evidence.

Savage made many contributions to the development of Bayesian statistics, of which the most significant are these: He proved a "washing-out" theorem that shows how, under fairly unrestrictive conditions, Bayesian agents with diverse prior probabilities will eventually converge to the same posterior given a sufficiently long run of shared observations. In a highly influential paper, written with Ward Edwards and Harold Lindeman (1963), he established the principle of stable estimation, which specifies conditions under which the value of a posterior probability will be independent of its prior. In one of his last papers, he developed an elegant general method for eliciting personal probabilities using proper scoring rules (1971). Savage died November 1, 1971, in New Haven, Connecticut, after having made lasting and seminal contributions to statistics, decision theory, psychology, and economics.

**See also** Bayes, Bayes' Theorem, Bayesian Approach to Philosophy of Science; Decision Theory; Statistics, Foundations of.

### **Bibliography**

- de Finetti, Bruno. "La prévision: Ses lois logiques, ses sources subjectives." *Annals de l'Institut Henri Poincaré* 7 (1937): 1–68. Translated as "Foresight: Its Logical Laws, Its Subjective Sources." In *Studies in Subjective Probability*, eds. Henry Kyburg and Howard Smokler. New York: John Wiley, 1964.
- Edwards, Ward, Harold Lindeman, and Leonard J. Savage. "Bayesian Statistical Inference for Psychological Research." *Psychological Review* 70 (1963): 193–242.
- Lindley, Dennis. "L. J. Savage—His Work in Probability and Statistics." *The Annals of Statistics* 8 (1980): 1–24.
- Joyce, James M. *The Foundations of Causal Decision Theory*. Cambridge, U.K.: Cambridge University Press, 1999.
- Ramsey, Frank P. "Truth and Probability." In *The Foundations of Mathematics and Other Logical Essays*, edited by Richard Braithwaite. London: Kegan Paul, 1931.

### WORKS BY SAVAGE

- The Foundations of Statistics*. New York: Wiley, 1954. 2nd revised edition, New York: Dover, 1972.
- "The Elicitation of Personal Probabilities and Expectations." *Journal of the American Statistical Association* 66 (1971): 783–801.

**James M. Joyce (2005)**

## SAVIGNY, FRIEDRICH KARL VON

(1779–1861)

Friedrich Karl von Savigny, the founder of historical jurisprudence, was born in Frankfurt, Germany, into a family that had moved there from Lorraine. Left an orphan at thirteen, Savigny was brought up by a friend who educated him in ways that recall the experience of young John Stuart Mill. At seventeen Savigny entered the University of Marburg; after studying at other universities, he returned to Marburg for his doctor's degree in 1800 and began a long, influential, and distinguished teaching career. At the age of twenty-four he published *Das Recht des Besitzes* (The Right of Possession; Giessen, 1804), and in the following year he began to tour libraries in search of manuscripts for his historical work. In 1810 he accepted a teaching post at the newly founded University of Berlin, which he helped organize and where he became rector. He did much to raise the standards of German universities and to help them achieve a dominant position in the world of scholarship. While teaching, writing, and assisting in the administration of the university until 1842, he also performed judicial tasks, and from 1842 to 1848 he was chancellor of Prussia.

In his stress on continuity and tradition Savigny may have been influenced by Edmund Burke, and in his understanding of the methods and aims of historical research he may have been influenced by Barthold Georg Niebuhr, who also took part in the founding of the University of Berlin and was an admirer of Roman institutions.

Savigny's two *magna opera* were the *Geschichte des römischen Rechts in Mittelalter* (7 vols., Heidelberg, 1815–1834) and the *System des heutigen römischen Rechts* (8 vols., Berlin, 1840–1849). In 1850 his miscellaneous writings, *Vermischte Schriften*, were published at Berlin in five volumes, and in 1851 and 1853 his two-volume work *Das Obligationenrecht als heute römischen Rechts* was published. He was cofounder, in 1815, of the *Zeitschrift für geschichtlichen Rechtswissenschaft*. His massive work on Roman law in the Middle Ages became the source of subjects for countless historical monographs. His students, and their students in turn, dominated historical and legal scholarship and teaching for several generations, and he was universally acknowledged as one of the most influential thinkers and scholars of the nineteenth century.

The main thrust of Savigny's jurisprudential thought, however, is not found in his monumental his-

torical treatises but in a polemical tract published at Tübingen in 1814, *Vom Beruf unserer Zeit für Gesetzgebung und Rechtswissenschaft*. This pamphlet was in rebuttal to A. F. J. Thibaut's *Civilistische Abhandlungen* (Heidelberg, 1814), in which a plan for a single code of laws for all German states was urged.

Savigny argued that law has no abstract origin in nature or mind but is organically connected with the people of a nation and is an expression of its *Volksgeist*, or collective genius. Fundamentally, law is formed by custom and popular faith, "by internal, silently operating powers, not by the arbitrary will of a lawgiver." The "real law" is always "the proper will of the people." Like language and manners, law has movement and development; it grows with a people and dies with it.

In earliest historical times, Savigny claimed, law was no more separable from a people than was its language or its manners. Rights and duties were created and extinguished by symbolic acts, which were the "true grammar" of law in this period. As social existence became more complex and sophisticated, law came to be expressed in abstract forms; jurists became a professional class, and law perfected its language and took a scientific direction. Instead of existing in the consciousness of the people, it now existed in the consciousness of the jurists, who became the representatives of the community, the voice of its *Volksgeist*. Now the law had a twofold existence: the "political" element, or the connection of the law with the general existence of the people, and the "technical" element, or the abstract and scientific existence of the law. From this it follows that the jurist needs a twofold spirit: the historical sense, with which to seize "the peculiarities of every age and every form of law," and the systematic sense, with which to see "every notion and every rule in lively connection and cooperation with the whole" legal order. Through these senses the jurist will acquire mastery over a body of law, obtain for that law a thorough grounding in history, and discover its organic principle. He will be able to separate that which still has life from that which is lifeless "and only belongs to history," and in this way he will arrive at a truly national law—a "living customary law."

Savigny's views contributed in varying degrees to a number of significant results: (1) They helped bring to an end the dominant natural law philosophy that looked to pure reason as the source of law. (2) They delayed the movement for codified legal systems that had started with the Napoleonic codes. (3) They established the historical school of jurisprudence. (4) They laid the basis for the sociological school of legal thought. (5) They retarded the

development and acceptance of legislation as a source of law. (6) They contributed to an exaggerated stress on nationalism and to a disparagement of the idea of a common law of humankind as an expression of *Menschengeist*. Perhaps Savigny's most enduring influence is to be found in his idea that law must not be isolated into an autonomous science but must be treated as an aspect of social life, development, and order—as a social, historically conditioned phenomenon.

**See also** Burke, Edmund; Historical School of Jurisprudence; Mill, John Stuart; Philosophy of Law, History of; Philosophy of Law, Problems of.

### Bibliography

#### PRIMARY WORKS

*Of the Vocation of Our Age for Legislation and Jurisprudence*. New York: Arno Press, 1975.

*System of the Modern Roman Law*. Translated by William Holloway. Westport, CT: Hyperion Press, 1979.

#### SECONDARY SOURCES

Kantorowicz, Hermann. "Savigny and the Historical School of Law." *Law Quarterly Review* 53 (1937): 326ff.

Montmorency, J. E. G. "Friedrich Carl von Savigny." In *Great Jurists of the World*, edited by J. Macdonell and E. Manson. Boston: Little Brown, 1914.

Stoll, Adolf. *Friedrich Karl von Savigny*, 3 vols. Berlin, 1927–1939.

Wolf, Erik. *Grosse Rechtsdenker der deutschen Geistesgeschichte*. Tübingen, 1951; 4th ed., 1963. Ch. 12.

Zwilmeyer, Franz. *Die Rechtslehre Savignys*. Leipzig, 1929. Bibliography on pp. ix–xii.

**Milton R. Konvitz (1967)**

*Bibliography updated by Philip Reed (2005)*

### SCEPTICISM

See *Skepticism, History of*

### SCHELER, MAX

(1874–1928)

A pioneering German phenomenologist, ethicist, and social philosopher, Max Scheler was born in Munich in 1874. His father was Lutheran, his mother was Jewish; Scheler himself, ever independent, embraced Catholicism. After studying with Wilhelm Dilthey and Georg Simmel, he earned his doctorate in 1897 under Rudolf Eucken in Jena, where he taught until 1906. From 1907 he taught in Munich, where he met Franz Brentano and sev-

eral disciples of Edmund Husserl, the father of the phenomenological movement. He soon became acquainted with a growing circle of phenomenologists from Munich and Göttingen, including Moritz Geiger (1880–1937), Dietrich von Hildebrand (1889–1977), Alexander Pfänder, Adolf Reinach (1883–1917), Edith Stein, and others. But as early as 1901, when he first met Husserl, Scheler had already taken an independent phenomenological direction of his own.

In 1910 Scheler lost his post in Munich after a divorce alienated him from the Catholic university administration. In 1912, he married Märit Furtwängler, sister of the noted conductor. From 1910 to 1919, he freelanced as an independent scholar, publishing a prolific number of works, particularly on ethics, but also on political issues of the day, including war, capitalism, feminism, the psychology of resentment, and various social issues. He served on diplomatic missions to Switzerland and the Netherlands. After World War I, he actively promoted the causes of international reconciliation, moral renewal, pacifism, and European reunification based on ideals of Christian socialism. It was not until 1919 that Scheler received a full professorship, in Cologne, where his focus turned to religion, anthropology, metaphysics, and sociology of knowledge. By 1922 he had fallen away from Catholicism in favor of a pantheistic conception of divine self-realization in history. He died on the eve of assuming his final post in Frankfurt in 1928, after repeatedly warning against the rise of German Nazism and Italian Fascism. His writings were suppressed by the Nazis in Germany from 1933 to 1945.

Scheler's impact on the phenomenological movement was considerable, despite ambivalent relationships with Husserl and Martin Heidegger. Many prominent thinkers have acknowledged their debt to him, including Jean-Paul Sartre, Maurice Merleau-Ponty, Gabriel Marcel, Nicolai Hartmann, Roman Ingarden, Hans-Georg Gadamer, Martin Buber, and José Ortega y Gasset. Pope John Paul II wrote a doctoral dissertation on him. Scholars in the Spanish-speaking world, Japan, and Russia were well acquainted with Scheler long before he was known in the English-speaking world.

Scheler's most important phenomenological works were published during his prolific middle period. These include *Formalism in Ethics and Non-Formal Ethics of Values* (1913–1916), his seminal critique of Immanuel Kant's ethics and outline of his own phenomenological ethics based on a theory of values. His 1916 essay *Ordo Amoris* develops his Pascalian conception of a faculty of cognitive feeling independent of reason, which apprehends a hier-

archical array of values in its pure incontrovertible immediacy. Between 1912 and 1913 he also published phenomenological studies of sympathy, love, and hate in *The Nature of Sympathy*, and a study of resentment and impotence in modern bourgeois morality in *Ressentiment* (1994 [1964])—a brilliant transmutation of Friedrich Nietzsche's claim that Judeo-Christian morality stems from resentment, eliciting Ernst Troeltsch's famous characterization of Scheler as “the Catholic Nietzsche.”

While initially collaborating with Husserl, Scheler criticized Husserl's “Cartesianism” and for giving inordinate primacy to reason. By contrast, Scheler insisted on the primacy of feeling and its independence from reason in apprehending values, which he considered the primordial phenomena of consciousness. Scheler did not use Husserl's terms *noesis* and *noema* to distinguish the act of thinking from the object of thought, yet he recognized that this polarity within consciousness, first investigated by Brentano, allows for two approaches in investigation. Thus he distinguished *act-phenomenology* from *phenomenology of facts*, the former focusing on persons as the source of the unifying intention animating acts, the latter analyzing three types of facts—natural, scientific, and phenomenological.

The preeminent phenomenological facts overlooked by Kant, according to Scheler, are values. Kant rightly denies that moral obligation can be defined by reference to empirical objects of desire without subordinating it to the relativizing contingencies of particular whims, ends, and purposes. But he fails to discern the distinctive nature of values as pure qualities or essences, distinct from empirical entities or objects of desire that might serve as their bearers. Just as colors can be conceived independently of any colored surfaces or bearers, values can be intuited as pure, independent essences. Furthermore, values exhibit an objective hierarchical ranking, furnishing a material basis for ethics, in contrast to Kant's empty formalism. Accordingly, Scheler distinguishes four basic ranks of values. From highest to lowest, these include the (1) *religious*, such as the sacred and profane; (2) *cultural*, such as the true, right, and beautiful; (3) *vital*, such as the noble and common; and (4) *sensory*, such as the pleasant and painful. Scheler's criteria for this classification are reminiscent of Jeremy Bentham's hedonic calculus, including relative duration, depth of satisfaction, and so forth. He also held that this ranking reflects an a priori “logic of preference.”

As in teleological theories generally, Scheler defines moral values in terms of the nonmoral value realized or intended through an act. Accordingly, moral good is

achieved as a by-product of realizing or intending a positive or comparatively higher nonmoral value, such as sacrificing the lower value of physical comfort for the higher value of one's children's education. His ethic, unlike Kant's, is based not on "blind duty," but on positive insight into the nature of values.

Scheler is unabashedly objectivist and absolutist in his value theory, but acknowledges the relativity of actual value judgments among societies and individuals. Someone suffering a pathological urge to sacrifice does not have the same obligation to be selfless as the self-centered egoist. Differences of cultural ethos are also significant. Recognition of such relativities inform Scheler's theories of virtue, conscience, and obligation, as well as his concepts of types of *exemplary acts* and *exemplary persons*—such as saints, geniuses, and heroes—that he proposes as vehicles for moral education. Yet he steadfastly maintains that such relativities do not undermine the absolute objectivity of values themselves.

*See also* Phenomenology.

### **Bibliography**

#### WORKS BY SCHELER

*Ressentiment*, edited with introduction by Lewis A. Coser. Translated by William W. Holdheim. New York: Free Press of Glencoe, 1961; Milwaukee, WI: Marquette University Press, 1994.

*Formalism in Ethics and Non-Formal Ethics of Values*. Translated by Manfred S. Frings and Roger L. Funk. Evanston, IL: Northwestern University Press, 1973.

*Max Scheler on Feeling, Knowing, and Valuing*, edited by Harold J. Bershady. Chicago: University of Chicago Press, 1992. Contains a listing of English translations.

#### SECONDARY WORKS

Blosser, Philip. *Scheler's Critique of Kant's Ethics*. Athens: Ohio University Press, 1995.

Frings, Manfred S. *The Mind of Max Scheler: The First Comprehensive Guide Based on the Complete Works*. Milwaukee, WI: Marquette University Press, 1997.

Kelly, Eugene. *Structure and Direction: Studies in the Phenomenological Philosophy of Max Scheler*. Boston: Kluwer, 1977.

Nota, John H. *Max Scheler: The Man and His Work*. Translated by Theodore Plantinga and John H. Nota. Chicago: Franciscan Herald Press, 1983.

Spader, Peter H. *Scheler's Ethical Personalism: Its Logic, Development, and Promise*. New York: Fordham University Press, 2002.

*Philip Blosser (2005)*

## SCHELLING, FRIEDRICH WILHELM JOSEPH VON (1775–1854)

Friedrich Wilhelm Joseph von Schelling, the German idealist philosopher, was born at Leonberg in Württemberg, the son of a learned Lutheran pastor, Joseph Friedrich Schelling. From his earliest years, he was destined by his family for the ministry. He was educated at the cloister school of Bebenhausen and, from 1790 to 1792, at the theological seminary at Tübingen. There he became friendly with two older students who were to play significant roles in his own life, as well as in cultural history: G. W. F. Hegel and J. C. F. Hölderlin, the great romantic poet. The three young men were keen partisans of the French Revolution, and they also enthusiastically discussed the ideas of the philosophers, especially Benedict de Spinoza, Immanuel Kant, and Johann Gottlieb Fichte.

For several years Schelling held a position as tutor of the sons of a noble family. Then, in 1798, at the unusually young age of twenty-three, he was called to a professorship at Jena. There the famous Fichte, the leading philosopher in Germany at the time and the idol of Schelling's youth, became his colleague and friend. In 1802 and 1803 Schelling and Hegel jointly edited the *Kritisches Journal der Philosophie*. At that time, though Hegel was five years older than Schelling, he was generally considered to be Schelling's disciple, and Hegel's first book was a comparison of Fichte's and Schelling's philosophies.

In nearby Weimar, Johann Wolfgang von Goethe and Friedrich Schiller were at the peak of their careers. Schelling met them both and became friendly with Goethe. Jena was now the center of German romanticism, and the ideas and personalities of this movement made a profound and lasting impression on Schelling. The romantic movement was, of course, also influenced by his philosophy. In its stress on the importance of the individual and the supreme value of art, and in its antirationalism, organicism, and vitalism, Schelling's transcendental idealism is the epitome of German romantic philosophies.

His friends among the romantics included Ludwig Tieck, who interested Schelling in folklore and mythology; the brilliant young poet Novalis; and August and Friedrich von Schlegel, whose translations of William Shakespeare made the English playwright one of the main shaping forces of German literature. Schelling was particularly intimate with August and his charming, intellectually gifted wife Caroline. Soon he became infor-

mally engaged to Auguste Böhmer, the sixteen-year-old daughter of Caroline by a previous marriage, but she died in 1800 before they could marry. It was rumored at the time that Schelling's amateur medical attentions contributed to her death. Certainly he was impetuous and self-confident to a point that some felt bordered on irresponsibility. This was a personal pattern common among the romantics, who sometimes defended themselves with the words of Schelling, "The beginning and end of all philosophy is—freedom."

In 1803 Caroline divorced August Schlegel and married Schelling. In keeping with the romantic creed, the three remained friends. It seems to have been an ideal marriage in every way. Schelling produced his most successful works during these years, and when Caroline died in 1809 he was grief-stricken; from then on he seemed unable to put his ideas together in a way that satisfied him. He never published another book as long as he lived, though he continued to write and lecture for many years. In 1812 he married Pauline Gotter, a friend of Caroline's.

From 1803 to 1806 Schelling taught philosophy at the new University of Würzburg, and in 1806 he was called to Munich as an associate of the Academy of Sciences and as secretary of the Academy of Arts. He later became secretary of the philosophical section of the Academy of Sciences. These positions were government sinecures that afforded him abundant leisure and also allowed him to lecture at Stuttgart and, from 1820 to 1827, at Erlangen. In 1827 he became a professor at Munich. In 1841 the Prussian authorities, in the hope that he would serve as a counterbalance to the powerful influence of the radical Young Hegelians, appointed him to the position of Prussian privy councilor and member of the Berlin Academy, and he lectured for the next five years at the University of Berlin. He died at the age of seventy-nine at Bad Ragaz, Switzerland.

Of all the major German philosophers, Schelling is the least known in the English-speaking world. His name is familiar as the historic link connecting Kant and Fichte with Hegel, but this description fits only his earlier work. Through his personal association with some of the German romantic writers and his doctrinal influence on the entire German romantic school, as well as through the direct influence of his aesthetics on Samuel Taylor Coleridge and, through Coleridge, his indirect influence on other English poets of the period, he is also known as the philosopher of romanticism. In his last phase, which was partly a conscious reaction to Hegel, he anticipated some of the central ideas of the existentialists, and for this

reason there has been a revival of interest in his later writings.

The development of Schelling's philosophy can be conveniently divided into four stages—subjective idealism, the philosophy of nature, the philosophy of identity, and the philosophy of the opposition of the negative and the positive. The stages are logically connected with one another, but also are clearly separate, so much so that their author was often accused of inconsistency. For example, Hegel wrote, "Schelling carried on his philosophic education before the public and signaled each fresh stage of his advance with a new treatise."

### SUBJECTIVE IDEALISM

In the first stage Schelling was gradually working himself free from Fichte's subjective idealism to an independent position of his own. The major works of this phase were *Vom Ich als Prinzip der Philosophie, oder über das Unbedingte im menschlichen Wissen* (Tübingen, 1795), in which he posited the ego as the supreme, unconditioned element in human knowledge, and *Philosophische Briefe über Dogmatismus und Kritizismus* (in *Philosophisches Journal*, 1796), in which he compared Spinoza and Fichte. There is little that is original in these works other than the style and the tone. However, Schelling's style is important because its eloquence, its sense of emotional urgency, and its relative freedom from technical jargon—a rare trait in the writings of German idealists—all point to his affinity with the romantic movement and his unique philosophic stress on the importance of aesthetics.

### PHILOSOPHY OF NATURE

The second stage, the philosophy of nature, was the most famous and the most influential of Schelling's philosophies and remained so until recent years. The first important work of this stage was *Ideen zu einer Philosophie der Natur* (Leipzig, 1797). Against Fichte's conception of the world as the construction of the ego, Schelling now insisted that the world of nature is just as real and just as important as the world of the ego. In fact, it is nature, the objective, that gives to consciousness what consciousness reproduces anew. Originally, consciousness and nature are one and infinite; but consciousness limits itself and presents itself to itself as finite, as different from nature. The essence of the ego is spirit, and the essence of nature is matter, but the essence of matter is force; that is, attraction and repulsion. In force, Schelling finds the common ground of nature and ego. As attraction it is objective, it is nature, it is matter; as repulsion it is subjective, it is ego, it is spirit. This duality also governs human perception:

As attraction to the self, force governs the streaming of the outer world into the inner world of sensation, and this internal experience of movement constitutes the a priori basis of time; as repulsion, pushing out into the world, force constitutes the a priori basis of space.

**PHYSICAL SCIENCES.** In *Von der Weltseele* (Hamburg, 1798) Schelling dealt with the philosophic problems of the physical sciences. He believed that the fundamental aim of the sciences was the interpretation of nature as a unity, and therefore the proper study of all sciences was force. He tried to show that mechanical, chemical, electrical, and vital forces were all different manifestations of the same underlying force. In the following year, in *Erster Entwurf eines Systems der Naturphilosophie* (Jena and Leipzig, 1799) and in *Einleitung zu dem Entwurf eines Systems der Naturphilosophie oder über den Begriff der spekulativen Physik* (Jena and Leipzig, 1799), he depicted this force as “pure activity.” He saw nature as an infinite self-activity, realizing itself in finite matter but forever unexhausted, forever short of completely realizing itself. He felt that he had thus found a parallel in the physical universe for Kant’s idea of the moral universe as practical reason forever striving toward an unattainable ideal. He further developed this phase of his thought in “Allgemeine Deduktion des dynamischen Prozesses” (in *Zeitschrift für spekulative Physik*, Vol. 1, 1800); *Über den wahren Begriff der Naturphilosophie; Darstellung meines Systems der Philosophie* (Jena and Leipzig, 1801); and *Bruno, oder über das göttliche und natürliche Prinzip der Dinge* (Berlin, 1802).

**KNOWLEDGE.** In the *System des transzendentalen Idealismus* (Tübingen, 1800), his most systematic and mature statement, Schelling applied to the philosophy of nature the insights gained from the Kantian and Fichtean philosophy of knowledge. His technique for deriving the world of objects from the world of the ego was to turn consciousness upon itself as the only object of which we have immediate firsthand knowledge. Thus, he found that when we abstract from all objects of knowledge, both within ourselves and in the outside world, we arrive at the pure activity of abstracting, which is pure self-activity. Seen in this light, the consciousness of the not-self is the limit of self-activity, just as the things-in-themselves are at the limits of knowledge in *The Critique of Pure Reason*.

On this foundation, Schelling built a theory of three stages of knowledge, which he described as progressing from sensation to perception, from perception to reflection, and from reflection to will. At first, consciousness of a limit, of the not-self, is felt as a sensation. The limit,

where the sensation is felt, is the meeting place of self-consciousness pushing outward and the force of the consciousness of external objects streaming inward. Therefore, all sensation is a feeling of myself as limited. Here we become aware of gravity, of the force of the real objective world in space, and also of intensity, which is the immediate consciousness of the self and its own activity in time. From the perception of the outside world comes reflection, and from reflection on the internal world comes will.

In this way Schelling felt that he had established links among Kant’s categories, schemata, and objects of perception. Aside from the technical question of the correctness of this linkage—certainly Kant would have disputed it—it has great historical importance, because this is perhaps the only area in which Schelling decisively influenced the fully matured philosophy of Hegel, who used this reasoning to connect the dialectic of thesis-antithesis-synthesis with Kant’s triadic formulation, though the dialectic itself was borrowed by Hegel from Fichte.

Schelling argued that the separation of knowledge from its object occurs only in abstraction. In reality, concepts have no existence apart from their objects, since knowledge is the meeting of objects and self. Therefore, the self is not merely one of the objects of knowledge; it is the condition of all knowledge. And since the essence of the self is pure self-activity, knowledge ultimately derives from willing, which is the action of the self.

**OTHER MINDS.** Schelling now asks two fundamental questions. How do I know there are other intelligences? And how can they act on me? He answers that our consciousness of limitations implies the existence of other selves that act as limiting factors. (Here he takes issue with Kant’s teaching that intelligence is limited by something not itself.) But the other selves can act on me only indirectly, through my representation of their acts. Their action does not compel mine, but limits it; and such limitation is compatible with my freedom. It is the community of interacting intelligences that constitutes the historical life of man. And while nature exists when not perceived by me, it exists then only because it is perceived by other human beings. Objectivity is intersubjectivity.

**WILL AND IMAGINATION.** Although perception is necessary and limited, will is free and unlimited. The imagination and its ideas mediate between perception and will. As opposed to the conceptions of the understanding, which are finite, the ideas of the imagination are both

finite and infinite. An idea's relation to its object is finite, but the activity of the imagination in this relation is infinite. Each idea is subsumed under an ideal, as conceptions are subsumed under their schemata in Kant. The function of the will is to idealize the imagination's ideas. The contradiction thus engendered gives rise to impulse, defined as the desire to restore destroyed identity. Through impulse, there is constant realization of ideals, but the ideas of the imagination are constantly striven after and never attained.

**WILL AND KNOWLEDGE.** The distinction between will and intelligence thus is relative, not absolute. From a higher point of view, they are identical. In intelligence, the I that acts and the I that knows are one. The acting I is an object for itself, while the knowing I merely perceives other objects. In action there is no transition from the world of nature to the world of mind, for the subject has become an object to itself. Any change in the outer world is received as a perception, but every action causes such a change; therefore action is perception. (Here, as elsewhere, Schelling anticipates Gestalt psychology.) Self-determination is the primary condition of all consciousness.

**JUSTICE.** The object of impulse, which always acts to restore the lost identity of the self and the world, is happiness. But an impulse that transcends its proper limits acts against itself and must be prevented by a sanction not found in nature—a sanction of the will. This sanction of the will is thus the basis of justice, and the law of justice is a second nature that our will sets above the first nature.

**THE NATURE OF HISTORY.** The process of history is the gradual realization of law; history can be described as the development of human freedom, as an eternal progress toward the perfect state—a sovereign world federation of all sovereign states—in which all men would be citizens. Thus, history is the realization of freedom through necessity. There is an absolute identity between freedom and necessity, but this identity is forever unconscious, never the object of knowledge but always the object of faith. God is neither personal nor objective, but the revelation of the divine in man. This revelation is never complete. History is a drama in which human beings are not merely the actors, but also the authors.

**ART AND AESTHETICS.** If history is a drama for Schelling, nature is a work of art. Like Kant in *The Critique of Judgment*, Schelling believed that organisms and works of art are alike in that they can be properly under-

stood only teleologically; that is, as entities in which the parts serve the whole and the whole is itself purposive. The main difference between art and organisms, according to Schelling, is that in organisms the activity of the organizing intelligence lies hidden or unconscious, manifest only in the product—the organism itself; but in the work of art the productive activity is conscious whereas the product, the true art work, is unconscious and infinite. The artist never fully understands his art. The purpose of art is neither utility, nor pleasure, nor morality, nor knowledge, but beauty—the realization of the infinite in the finite.

In his aesthetics, which is elaborated in the *System des transzendentalen Idealismus* and his lectures on the philosophy of art, *Über das Verhältniss der bildenden Künste zu der Natur* (Munich, 1807), Schelling is at his most personal, his most impassioned, his most characteristic, and his most original. He held that in art, intelligence for the first time becomes completely self-conscious. In philosophy, it is abstract and limited in the expression of its potential infinity. But in art, which is completely free from abstraction in this sense, intelligence fully realizes its infinite nature. (It is pertinent that Hans Arp, the abstract artist, has written that the works usually called “abstractions” are more accurately referred to as “concretions.”) Thus art is the goal toward which all intelligence moves. Art is the true philosophy, because in it nature and history are forever reconciled; but the artist is not therefore a philosopher, since he often lacks a theoretical understanding of his own creation. The theoretical intelligence merely contemplates the world, and the practical intelligence merely orders it; but the aesthetic intelligence creates the world.

## PHILOSOPHY OF IDENTITY

The third stage of Schelling's thought was the philosophy of identity, first expounded at length in *Vorlesungen über die Methode des akademischen Studiums* (Tübingen, 1803), appropriately written in Spinoza's geometric mode. Here Schelling said that the philosophy of nature and the philosophy of knowledge, taken together, constitute only half the truth and need to be completed by the other half, which unites nature and knowledge in an undifferentiated identity. The production of reality does not rest on the opposition of intelligence and nature, subject and object, but in the identity of all reality as it rises from the absolute. The absolute identity of nature and intelligence is found in their common neutral source, reason. Reason is one and infinite, embracing things-in-themselves and knowledge of things. In reason there is no



object, no subject, no space, no time. Its supreme law is the law of identity,  $A = A$ , which is true regardless of all spatial or temporal considerations. In the formula  $A = A$ , the distinction between subject and object is formal and relative. Subject and object here concern only the form, and are indifferent as to essence. It was this phase of Schelling's thought that Hegel wittily called "the night in which ... all cows are black."

**PANTHEISM.** The philosophy of identity was a kind of pantheism, but it stressed the aliveness of nature in contradistinction to Spinoza's dead, materialistic, deterministic pantheism. Although Spinoza's influence is evident, it is filtered through the vitalistic interpretations of Johann Gottfried Herder and Goethe and tempered by the parallel influence of Giordano Bruno's vitalistic pantheism. Schelling believed that life was the basis of the inorganic world, and not vice versa. Nature is inseparable from God, but distinguishable from him. God is not to be comprehended rationally, because his essence is will and he can be apprehended only through the will, in action. For the most part, Schelling's thought here draws from Jakob Boehme, and reintroduces Protestant mysticism into the mainstream of Western philosophy.

**GOD AND EVIL.** In *Philosophische Untersuchungen über das Wesen der menschlichen Freiheit* (Landshut, 1809; translated as *Of Human Freedom*, Chicago, 1936) Schelling, like Boehme, distinguishes between God as ground of being and God as perfection. Evil is explained as the ground eliciting the self-will of man in order to awaken him to the distinction between good and evil, which originally were united in one identity. Thus, evil is a necessary stage in the progress toward the total realization of good. Imperfection in being is perfection in the process of becoming. There is a dark ground or negative principle in God, but it exists so that he can become separate from it as a personality.

### POSITIVE PHILOSOPHY

After 1809, the year of his first wife's death, Schelling made the given situation of existence his predominant concern. This final existentialist phase of his philosophy was first propounded in *Die Weltalter* (written in 1811 but not published in Schelling's lifetime), consummated in his lectures at the University of Berlin, and saved for posterity in three volumes, *Einleitung in die Mythologie*, *Philosophie der Mythologie*, and *Philosophie der Offenbarung*, which were published posthumously in the *Sämtliche Werke*. In these works he sought to erect a positive philosophy based on the evolution of the divine

principle in human history, especially in myths and religions, which he felt opposed and thus completed his own earlier, negative, merely rational philosophy. However, rather than representing a sharp break with his past, this last phase can be considered as the flowering of tendencies he showed as early as 1795, when he wrote, "The main function of all philosophy is the solution of the problem of the existence of the world." It is significant that while the prolific and influential writings of his first three periods were crowded into fourteen brief years, from 1795 to 1809, his last period, during which his rate of production slowed and his influence waned, lasted from 1809 to his death in 1854.

**GOD.** The root of existence is now found in nonbeing, in God as the ungrounded, the abyss, the eternal nothing. Only against the ungrounded can the ground arise, because nothing can become evident without resistance. Thus God is "eternal contrariety," forever alienating himself from himself. This alienation creates the possibility of the fall. As only the Absolute is real, finite things, which are not real, can exist only in a removal, in a fall from reality. The Absolute creates its own counterpart, freedom, which is both the cause of the fall and the last trace of divinity things bear after the fall. Because of this progression through opposites, Schelling called this fourth phase of his thought the opposition of negative and positive philosophy.

**MAN.** As the creature in whom the fall, and the state of things before the fall, both rise for the first time into consciousness, man is the crown of creation and the most interesting and rewarding object of philosophic attention. Man is free creative activity, the essence of the world. Thus, in his last phase, Schelling was led to a kind of philosophic anthropology, seeking for the essence of man in what he thought was his deepest activity, myth-making and religion. Despite the profoundly mystical flavor of his thought in this period, he still kept contact with his Kantian heritage. In *Philosophie der Mythologie* he explained mythology as a symbolic system of ideas with its own a priori structure as necessary for its functioning as, according to Kant, the a priori structure of the understanding is necessary for logical thought. Ernst Cassirer's neo-Kantian formulation of mythology as just such a conceptual structure owes a great deal to Schelling, a debt fully acknowledged in the second volume of *The Philosophy of Symbolic Forms*.

**RESEMBLANCE TO EXISTENTIALISM.** What has made this last phase of Schelling's thought most apposite to

modern existential philosophy is another question rising from his consideration of man's being in the world. As he put it, "Just he, man, impelled me to the final desperate question: Why is there anything at all? Why not nothing?" It is this question, described as "dreadful" rather than "desperate," that Martin Heidegger took for his central theme in *Being and Time*.

Schelling's resemblance to the modern existentialists is suggestive rather than substantive, but the suggestion is inescapable. Like them, he emphasized that philosophy must deal not only with the "what" of the world, which explains its nature, but also with the "that" of the world—the fact of its existence, of its being there. And like Søren Kierkegaard (who attended some of his lectures in Berlin but was not impressed), Friedrich Nietzsche, Heidegger, and Jean-Paul Sartre, Schelling tried to express the inexpressible pathos of existence in oracular utterances halfway between poetry and metaphysics, the quality of which can be conveyed only by quotation. The world and God have as common ground "the incomprehensible basis of reality." "Existence is self-affirmation." God is "the infinite affirmation of himself." The objective world is the unconscious poetry of the spirit creating itself. Finally, there is a striking formulation of the existential anxiety, which is also an anticipation of the psychoanalytic doctrine of resistance: "The philosopher who knows his calling is the physician who . . . seeks to heal with gentle, slow hand the deep wounds of human consciousness. The restoration is all the more difficult since most people do not *want* to be healed at all and, like unhappy patients, raise an unruly outcry if one even approaches their wounds."

So the problems posed by Schelling in the nineteenth century are still very much alive in the philosophic and literary world of today. At that time his main influence in England was in aesthetics, and his lectures on the philosophy of art were translated as *The Philosophy of Art* in 1845. The continuing, perhaps growing contemporary interest in him is demonstrated by the fact that the first translations into English of any of his books since then—significantly, both from his last, existentialist phase—were published in America in 1936 and 1942.

## Bibliography

### WORKS BY SCHELLING

- Grundlegung der positiven Philosophie. Münchener Vorlesungen WS 1832/33 und SS 1833*, edited by H. Fuhrmans. Turin: Bottega d'Erasmus, 1972.
- Einleitung in die Philosophie*, edited by W. E. Ehrhardt. Stuttgart: Fromman-Holzboog, 1989.

*System der Weltalter*, edited by S. Peetz. Frankfurt: Klostermann, 1990.

*Urfassung der Philosophie der Offenbarung*, edited by W. E. Ehrhardt. Hamburg: Meiner, 1992.

### Collected Works

*Sämtliche Werke*. 14 vols, edited by K. F. A. von Schelling. Stuttgart and Augsburg: Cotta, 1856–1861.

*Werke*. 8 vols, edited by M. Schröder. Munich, 1927–1956.

### Translations

*The Ages of the World*. Translated by F. de Wolfe Bolman. New York: Columbia University Press, 1942. Translation of *Über die Gottheiten von Samothrake*, Stuttgart and Tübingen, 1815, a part of *Die Weltalter*, which is mentioned in the text. *The Ages of the World* includes a long, informative introduction on Schelling's philosophy, especially its last phase.

*Of Human Freedom*. Translated by J. Gutmann. Chicago: Open Court, 1936. Translation of *Philosophische Untersuchungen über das Wesen der menschlichen Freiheit*.

*The Philosophy of Art: An Oration on the Relation between the Plastic Arts and Nature*. Translated by A. Johnson. London: Chapman, 1845. Translation of *Über das Verhältniss der bildenden Künste zu der Natur*.

*System of Transcendental Idealism*. Translated by P. Heath. Charlottesville: University Press of Virginia, 1978.

*The Unconditional in Human Knowledge: Four Essays, 1794–1796*. Translated by Fritz Martl. Lewisburg, PA: Bucknell University Press, 1980.

*Bruno, or On the Natural and the Divine Principles of Things*. Translated by Michale Vater. Albany: State University of New York Press, 1984.

*Ideas for a Philosophy of Nature: An Introduction to the Study of This Science*. Translated by E. E. Harris and Peter Heath. Cambridge, U.K.: Cambridge University Press, 1988.

*On the History of Modern Philosophy*. Translated by Andrew Bowie. Cambridge, U.K.: Cambridge University Press, 1994.

### WORKS ON SCHELLING

Beach, E. *The Potencies of God(s): Schelling's Philosophy of Mythology*. Albany: State University of New York Press, 1994.

Bowie, A. *Schelling and Modern European Philosophy*. London: Routledge, 1993.

Bréhier, Émile. *Schelling*. Paris: Alcan, 1912.

Cassirer, Ernst. *Das Erkenntnisproblem in der Philosophie und Wissenschaft der neueren Zeit*, Vol. III: *Die Nachkantischen Systeme*. Berlin: Cassirer, 1920. Includes a comprehensive, clear, and sympathetic account of Schelling's epistemology.

Copleston, Frederick C. *History of Philosophy*. Vol. VII. London, 1963. Includes the best account of Schelling's philosophy in English.

Courtine, J.-F. *Extase de la raison: Essais sur Schelling*. Paris: Galilée, 1990.

Esposito, J. *Schelling's Idealism and Philosophy of Nature*. Lewisburg, PA: Bucknell University Press, 1977.

Fischer, Kuno. *Geschichte der neueren Philosophie*. Vol. VII. 3rd ed. Heidelberg, 1902. Probably the best full-length study of Schelling's life and work.

Frank, M. *Der Unendliche Mangel am Sein: Schellings Hegelkritik und die Anfänge der Marxschen Dialektik*. Frankfurt: Suhrkamp, 1975.

- Hartmann, Nicolai. *Die Philosophie des deutschen Idealismus*. 2nd ed. Berlin, 1960. An exhaustive and authoritative work on German idealism.
- Heidegger, M. *Schelling's Treatise on the Essence of Human Freedom*. Athens: Ohio University Press, 1985.
- Hirsch, Eric D. *Wordsworth and Schelling*. New York, 1962. Reviews Schelling's aesthetics and its parallels in William Wordsworth.
- Jaspers, Karl. *Schelling: Grösse und Verhängnis*. Munich: Piper, 1955. A profound and searching analysis from an existentialist point of view.
- Kaspar, W. *Das Absolute in der Geschichte*. Mainz: Grünewald, 1965.
- Kroner, Richard. *Von Kant bis Hegel*. 2 vols. Tübingen: Mohr, 1921–1924.
- Marx, W. *The Philosophy of F. W. J. Schelling: History, System, and Freedom*. Bloomington: Indiana University Press, 1984.
- Noack, Ludwig. *Schelling und die Philosophie der Romantik*. Berlin: Mittler, 1859.
- Read, Herbert. *The True Voice of Feeling*. London: Faber and Faber, 1953. Concerned with Schelling's influence on Coleridge. Includes translation of a chapter on philosophy of art by Schelling.
- Schneeberger, G. *Schelling: Eine Bibliographie*. Bern, 1954. An extensive bibliography.
- Schulz, W. *Die Vollendung des deutschen Idealismus in der Spätphilosophie Schellings*. Pfullingen: Neske, 1975.
- Snow, D. E. *Schelling and the End of Idealism*. Albany: State University of New York Press, 1996.
- Tillich, P. *The Construction of Historical Religion in Schelling's Positive Philosophy*. Lewisburg, PA: Bucknell University Press, 1974.
- Tilliette, X. *L'Absolu et la philosophie: Essais sur Schelling*. Paris: Presses Universitaires de France, 1987.
- Watson, John. *Schelling's Transcendental Idealism*. Chicago: Griggs, 1882. A clear, careful, scholarly paraphrase of Schelling's *System des transcendentalen Idealismus*.
- White, A. *Schelling: An Introduction to the System of Freedom*. New Haven, CT: Yale University Press, 1983.

**Adam Margoshes (1967)**

*Bibliography updated by Thomas Nenon (2005)*

## SCHILLER, FERDINAND CANNING SCOTT (1864–1937)

Ferdinand Canning Scott Schiller, the British pragmatist philosopher, was born in Schleswig-Holstein and studied at Rugby and at Balliol College, Oxford. After teaching German at Eton, he returned to Oxford for his MA. In 1893 he went to Cornell University as an instructor and graduate student. In 1897, without receiving a doctorate, he returned to Corpus Christi College, Oxford, where he was successively assistant tutor, tutor, senior tutor, and fellow and where he received a DSc in 1906. He served as

treasurer of the Mind Association and president of the Aristotelian Society (1921), and he was elected a fellow of the British Academy in 1926. From 1926 on, Schiller spent part of each year at the University of Southern California as visiting lecturer and then as professor; in 1935 he moved there permanently.

### PRAGMATISM

Schiller's views, which he called at various times humanism, voluntarism, and personalism, as well as pragmatism, were strongly influenced by William James; and Schiller paid James great tribute, although he claimed to have arrived at his opinions independently. There was, however, an important difference of emphasis between them: James stressed the purposive aspect of thinking, and Schiller, the personal. James also accepted the independence of what is objectively given, whereas Schiller regarded all knowledge, even of "facts," as relatively subjective. Both Schiller and John Dewey were strongly influenced by G. W. F. Hegel and took the process of knowing as central to reality, but the influence of idealism was much stronger on Schiller than on Dewey. And whereas Schiller pursued the subjective and individual aspects of James's psychology, Dewey built upon its objective and social aspects. C. S. Peirce thought that Schiller's philosophy was intermediate between James's and his own.

Schiller's views may best be understood in terms of his opposition to the dominant absolute idealism of the British Hegelians, F. H. Bradley (Schiller's particular *bête noire*), J. M. E. McTaggart, Bernard Bosanquet, and T. H. Green. To Schiller the absolutism, monism, authoritarianism, rationalism, and intellectualism that these thinkers espoused ignored the basic insight of Protagoras that man is the measure of all things.

Schiller was convinced that all acts and all thoughts are irreducibly the products of individual human beings and therefore inescapably associated with the needs, desires, and purposes of humans. Such terms as *reality* and *truth* denote nothing complete and absolute; rather, they are intertwined with human intentions and deeds. Schiller emphasized the effective creativity of the human mind in organizing the universe of human experience and thus in making or remaking "reality." Man makes his truth along with his other values, Beauty and Goodness. Our axioms are never God-given but are human-made; they are not a priori verities but postulates, or working hypotheses, whose truth grows or diminishes within our experience. The logic we employ in gathering knowledge is dynamic and functional rather than eternally fixed. Our data are not "the given" but "the taken." Thus, in

Schiller's view, human activity is focal both to epistemology and to metaphysics, and there is genuine novelty in our growing universe and no theoretical limit to human freedom.

### MAKING REALITY

The absolute idealists maintained that reality is a seamless logical unity, not a mere disjointed plurality; that in the Absolute all separateness vanishes; that nothing finite, nothing that changes, is ever quite real, not even human personality; and that there is something makeshift, transitory, and unsatisfactory about the bits of matter we see, the individual acts we perform, and the private thoughts we think. But, Schiller pointed out, that is all that exists for us. An independent or absolute reality that does not enter into our experience, or explain our knowledge, is irrelevant to us. "Reality" for us is piecemeal, incomplete, and plastic. It is idle to ask "What is real?" Rather, the only question we can answer is "What can I know as real?"

The reality revealed by our actual active procedures of knowing is not rigid but malleable, not completed but evolving. Because it responds, at least to some extent, to our working and probing, it must somehow be not unrelated to our needs and purposes. The process of knowing, Schiller said, is "never one of bringing the mind into relation with a fundamentally alien reality, but always one of improving and extending an already existing system which we know." What we call real is that which, for our own reasons, we evaluate as important. It is the result of the kind of selection by which we reduce the chaos about us to order.

Schiller's critics found intolerable the thesis that we make reality. Bertrand Russell, for example, wrote, "Dr. Schiller says that the external world was first discovered by a low marine animal he calls 'Grumps,' who swallowed a bit of rock that disagreed with him, and argued that he would not have given himself such a pain, and therefore there must be an external world. One is tempted to think that ... many people ... had not yet made the disagreeable experience which Grumps made. Meanwhile, whatever accusations pragmatists may bring, I shall continue to protest that it was not I who made the world" ("Professor Dewey's 'Essays in Experimental Logic,'" *Journal of Philosophy* 16 [January 1919] 26).

Schiller found it hard to meet two particular objections to the theory of the making of reality: The world obviously preceded the existence of humans, and there are patent limits to human powers. In his later writings Schiller therefore reluctantly accepted the distinction between "finding" and "making" the real, although he

reiterated the meaninglessness of the "real-as-it-is-in-itself." He revived the Greek term *hule* to refer to the indeterminate, formless chaos, to whatever may be beyond man's ability to perceive or manipulate, to the raw malleable material of the cosmos.

Despite its drawbacks, the doctrine of the making of reality provided Schiller with the basis for certain important conclusions. In his view, it provided a perfect accommodation for Darwinian evolution; it supported a belief in the existence of genuinely new things and situations (always a problem for the absolute idealists because they regarded reality as a self-contained whole); it legitimized human progress; it provided a suitable conceptual scheme for the view, which Schiller ascribed to Albert Einstein and other scientists, that to posit "the real" independently of our sensations is to make an intellectual construction; and, most significantly, it was a firm foundation for man's freedom.

### OTHER METAPHYSICAL VIEWS

Schiller's other metaphysical views may be briefly stated. The function of philosophy, he thought, was to preserve the grand synoptic vision, to be an ultimate synthesis of the special sciences. Metaphysical systems, he held, are quasi ethical, or even aesthetic, in character; they reflect personality and temperament. Because the individual human person was an ontological ultimate for Schiller, he was a personalistic pluralist. He was also a hylozoist, asserting that all matter is more or less alive.

### TRUTH

Many theories of truth have been propounded through the centuries, but none has been entirely satisfactory. Schiller pointed out the shortcomings of some, particularly the correspondence and coherence theories. Pragmatists agree that no statement wears its truth like a badge; its truth can be determined only by what follows from it in the course of experience. Truth is only a potential, a valuation applied as the result of a procedure called verifying, or making true. Truth is relative to the evidence and to the purpose of the investigator; no degree of verification will ever establish the absolute truth of a statement. Schiller held that truth is personal and particular, dynamic and progressive, not eternal or absolute but the best solution found so far for any problem. That which thwarts or defeats the purpose of an inquiry we call false; that which furthers it we call true. "Truth is that manipulation of [objects] which turns out upon trial to be useful, primarily for any human end, but ultimately for that

perfect harmony of our whole life which forms our final aspiration" (*Humanism*, p. 61).

Nevertheless, Schiller thought the conversion of "The truth is useful" to "The useful is true" to be malicious. Therefore, in Chapter 8 of *Logic for Use* he distinguished seven kinds of truth claims. (1) A postulate is a statement that is "desirable if true," whose truth we try to establish. (2) "A fully verified postulate which serves as principle for a fully established science" and "rests securely on the solid mass of scientific fact it has been instrumental in eliciting" is an axiom. (3) A methodological assumption (determinism, for example) is any guiding principle that appears to be useful in analyzing the flux of events. (4) An assumption of limited usefulness, such as the use of Euclidean geometry in cartography, is a methodological fiction. Finally, truth claims may be, or are, made in (5) fictions, (6) jokes, and (7) lies. Lies are deliberately untrue but may be useful, as in propaganda.

Thus, Schiller held, to claim that all truths work for us in some way and that there is no useless knowledge is far from saying that whatever is useful is true. However, he was aware of difficulties concerning the status of past truth, the usefulness of some parts of pure mathematics, and such questions as whether truth is equivalent to survival value or to social acceptance.

## LOGIC

Since the true is what is true for us as seekers for it, Schiller deplored the divorce of logic from the empirical sciences and from psychology. He criticized traditional formal logic for having been a word game and for having been allied to metaphysics rather than to the empirical sciences and to psychology. For Schiller, as for Dewey, thought arises as an element in the solution of a problem. Thus the activity of reasoning has a biological matrix, and it is conditioned by such factors as interest, purpose, emotion, and satisfaction. Schiller was concerned with showing that meanings had been misunderstood and ignored by logic. Meanings, he pointed out, are acquired only in use; they are plastic and personal, and they occur only in contexts. Traditional logic regarded them as purely verbal and as fixed; it believed that one meaning corresponded to one form, and vice versa.

Schiller thought that logic had made the two mistakes of "etherealizing" and "depersonalizing" truth. In its search for formal validity, it had made three fatal abstractions; from actual thinking processes (psychology); from purpose, truth, or utility; and from meaning, matter, and context. In two books, *Formal Logic* (1912) and *Logic for Use* (1929), Schiller made an exhaustive study of formal

logic, including terms, propositions, definitions, the syllogism, and fallacies. He showed that, even on its own terms, logic was not free from ambiguity—how can there be novelty in the conclusion of a syllogism? What is the precise import of the copula in a proposition? Moreover, logic appealed at several crucial points to such psychological notions as the "necessity" of implication and the "certainty" or "self-evidence" of propositions. Schiller thought that logic should become a systematic evaluation of actual knowing, a study continuous with the sciences. His resolute experimentalism led him to assert, in "Axioms as Postulates" (1902), that even the laws of thought (identity, contradiction, excluded middle) are not principles of being or rules of logic but postulates.

## SCIENTIFIC METHOD

In analyzing the procedures of science, Schiller made several noteworthy contributions. He showed that the concept of "fact" is ambiguous. The "facts" of the scientist are the result of a process of selection, segregation, and evaluation; they are relative to the state of the science, the methods and instruments used, and the aims and bias of the scientist. They are also relative to the hypothesis used, to our own senses, to our memory, and to our words. Schiller also said, "The impossibility of 'breaking' a Law of Nature proves nothing but our determination to uphold a phraseology we have found convenient" (*Formal Logic*, p. 328).

## ETHICS AND RELIGION

Schiller carried his pragmatic approach into ethics and religion. There are no abstract values, he said, but only acts of personal valuation. Moral principles are not a priori presuppositions of right conduct; they are its results. The statements of religion are likewise postulates. (James spoke of the will to believe; Schiller, of the right to postulate.) God is a pervasive principle of goodness, not infinite but finite, struggling to develop; the actions of men therefore make a difference. Man's freedom is correlative to the postulate that man is responsible for his acts and is an agent in the full sense of the term. Schiller shared with James and Henri Bergson an interest in psychical research that stemmed from his desire to examine the methods of science at its periphery and from his postulate of immortality. Schiller was also keenly interested in eugenics. This led him to oppose democracy as a "sham" (*Problems of Belief*, p. 81) and to praise the British fascist Oswald Mosley. His social opinions were generally regarded by his philosophic supporters as a vagary.

Schiller was a prolific writer, a sprightly stylist, and a spirited polemicist who maintained a role of philosophic *enfant terrible* through hundreds of essays and books. He edited and wrote most of a parody of *Mind*, which he called *Mind!*—one of the rare examples of philosophic humor.

**See also** Humanism.

### **Bibliography**

Schiller's major books are two collections of essays, *Humanism: Philosophical Essays* (London and New York: Macmillan, 1903) and *Studies in Humanism* (London and New York: Macmillan, 1907). Two later collections are *Must Philosophers Disagree?* (London: Macmillan, 1934) and *Our Human Truths* (New York: Columbia University Press, 1939).

His major works in logic are *Formal Logic: A Scientific and Social Problem* (London: Macmillan, 1912) and *Logic for Use: An Introduction to the Voluntarist Theory of Knowledge* (London: G. Bell, 1929).

His essay "Axioms as Postulates" appeared in the anthology *Personal Idealism*, edited by Henry Sturt (London: Macmillan, 1902), pp. 47–133. An early, prepragmatic metaphysics, titled *Riddles of the Sphinx* and written under the pseudonym of A. Troglodyte (London and New York, 1891), was offered unsuccessfully to Cornell University as a PhD thesis. The essays "Scientific Discovery and Logical Proof" (1917) and "Hypothesis" (1921) were contributed to *Studies in the History and Method of Science*, edited by Charles J. Singer (Oxford: Clarendon Press, 1917–1921), and may be found in Vol. I, pp. 235–289, and Vol. II, pp. 414–446, respectively. Another work by Schiller is *Problems of Belief* (London: Hodder and Stoughton, 1924).

A critical study of Schiller's philosophy, with an exhaustive bibliography, is Reuben Abel, *The Pragmatic Humanism of F. C. S. Schiller* (New York: King's Crown Press, Columbia University, 1955).

See also Reuben Abel, ed., *Humanistic Pragmatism; The Philosophy of F. C. S. Schiller* (New York: Free Press, 1966); and *A Bibliography of the Works of F. C. S. Schiller, with an Introduction to Pragmatic Humanism* (San Diego, CA: San Diego State College Press, 1969).

**Reuben Abel (1967)**

*Bibliography updated by Michael J. Farmer (2005)*

## SCHILLER, FRIEDRICH (1759–1805)

Friedrich Schiller, a famed dramatist, poet, and essayist, was born in Marbach, a small town in southwest Germany, to Elisabeth Kodweiss and Johann Kaspar Schiller, a lieutenant in the army of the Duke of Württemberg. Though tutored in Latin at an early age by his local pastor to prepare him for theological studies, Schiller was

mandated by the duke to attend the duke's new military academy, Karlsschule. Schiller later related how his rebellion against the suffocating rigidity and isolation of Karlsschule paradoxically fostered his love of poetry. He remained at the school for eight years, focusing first on law, then on medicine. After his second medical dissertation, "On the Connection of the Animal Nature of Man with his Spiritual Nature," was accepted, he became a regimental physician in Stuttgart. There, he completed his first drama, *The Robbers*, the staging of which a year later (1782) in Mannheim brought him immediate acclaim and confirmation of his literary gifts. When the duke forbade him to write anything but medical treatises, Schiller fled Württemberg. For most of the rest of his life he would suffer considerable financial hardship and extremely poor health. Nevertheless, from 1782 to 1787 he managed to complete three plays (*Fiesco*, *Intrigue and Love*, and *Don Carlos*), to compose several poems (e.g., "Ode to Joy") and essays (e.g., "Theater Considered as a Moral Institution" and "Philosophical Letters"), and to found the journal *Rheinische Thalia*—all of which helped cement his reputation as a member of the *Sturm und Drang* (Storm and Stress) literary movement of the time.

While Schiller's literary output as a critic continued unabated in the ensuing years, his attention over the next decade (1787–1796) turned from the stage to the study of history and to an increasing preoccupation with philosophical treatments of morals and the arts. His *History of the Revolt of the Netherlands* (1787), which celebrated religious tolerance, won him a professorship (albeit unsalaried) in history at the University of Jena in 1789, and over the next two years he produced the enormously successful *History of the Thirty Years War*. His inaugural lecture, "What Does 'Universal History' Mean and to What End Is It Studied?" (1789) contains reflections, fairly conventional at the time, on history's progressive character. This progressive view of history collided, however, with a longing for a lost harmony that he thought art alone can provide (compare his nostalgic elegy of 1788, "The Gods of Greece," with his stirring, forward-looking call to his caste in the 1789 poem "The Artists").

This collision converged with a burgeoning interest in Immanuel Kant's moral and aesthetic writings. Following his marriage to Charlotte von Lengefeld in 1790 and an almost fatal bout with pneumonia a year later, Schiller was given the opportunity to pursue these interests in earnest thanks to a three-year pension provided by Prince Friedrich Christian von Schleswig-Holstein-Augustenburg of Copenhagen. Over the next four years Schiller composed several essays on aesthetics. The organ for

many of these essays was the journal *Die Horen*, founded by Schiller with the help of many of the leading figures in German letters at the time, among them Johann Wolfgang von Goethe and Wilhelm von Humboldt, with whom Schiller developed close friendships that had a lasting influence on his work.

Following this academic and philosophical interlude, and with Goethe's increasing encouragement, Schiller turned his attention back to the theater where he crowned his fame as a playwright with several historical plays: the *Wallenstein* trilogy (1798–1799), *Maria Stuart* (1800), *The Maid of Orleans* (1801), *The Bride of Messina* (1802), and *Wilhelm Tell* (1804).

### CRITICAL APPROPRIATION OF KANT'S PHILOSOPHY

Though philosophical concerns are apparent in Schiller's earliest publications, he makes his most influential philosophical contributions in essays composed between 1792 and 1796. The common feature of the first group of these essays is their critical engagement with Kant's philosophy. The aborted project of the "Kallias-Letters" (1793; published 1847) attempts in Kantian terms to establish something Kant declared impossible: "an objective concept of beauty" and, indeed, one that unites the realms of nature and freedom. In the "Kallias-Letters" Schiller accordingly construes beauty as "freedom in the appearance" of something, an appearance that is the natural or artistic, dynamic counterpart to moral autonomy. In "On Grace and Dignity" (1793) Schiller takes further aim at Kant's dualism, in particular, his account of an obligatoriness that is independent of grace ("the expression of a beautiful soul, where sense and reason harmonize").

Schiller's remarks provoke an exchange of letters and a public response in Kant's *Religion within the Limits of Reason Alone* (1793), where Kant suggests that any apparent disagreement can be resolved by distinguishing duty, the dignity of which is necessarily independent of grace, from virtue, which is not. Though Schiller accepts the suggestion in correspondence with Kant, he ultimately finds the distinction unpersuasive. Nevertheless, Schiller utilizes themes from Kant's aesthetics to develop a conception of tragedy in other essays from this period, notably, "On the Reason for Taking Pleasure in Tragic Subjects" and "On the Art of Tragedy" in 1792 and "On the Pathetic" in 1793. In particular, in Kant's notion of the dynamically sublime, the aesthetically pleasing displays of human beings' moral capacity to defy nature's otherwise all-powerful sway over them, Schiller finds the key to explaining the point of tragedy, though he invests art with

a purpose beyond the confines of Kant's aesthetics. As Schiller puts it in the opening lines of "On the Pathetic," "Portrayal of suffering—as mere suffering—is never the end of art, but as a means to this end it is of the utmost importance to art. The ultimate purpose of art is to depict what transcends the realm of the senses and the art of tragedy in particular accomplishes this by displaying morality's independence, its freedom, in the throes of passion, from nature's laws" (1993 [1793] p. 45).

### THE AESTHETIC LETTERS

Schiller's most influential work on aesthetics is *On the Aesthetic Education of Man in a Series of Letters* (1795). In this work (hereafter *Letters*) Schiller frames an argument for the necessity of an aesthetic education against the backdrop of a dire assessment of contemporary culture. Echoing Jean-Jacques Rousseau and anticipating Karl Marx, the assessment emphasizes the stupefying fragmentation and lifeless mechanism of society. Still, neither reason nor politics, Schiller argues, provides an answer to humanity's plight. The French Revolution had demonstrated only too well the failure of political reform without a moral transformation of the citizenry, that is, a transformation of individuals into citizens. As for reason, if it is the answer, Schiller asks, why in an "enlightened age" are we still barbarians? With art as the sole remaining alternative Schiller announces his central thesis, "If man is ever to solve the problem of politics in practice he will have to approach it through the problem of the aesthetic, because it is only through beauty that man makes his way to freedom" (1993 [Letter 2, 1795], p. 90). Though Schiller sometimes (e.g., Letter 14) ascribes freedom and morality solely to the rational side of human nature, the overriding sense of freedom at work in the *Letters* is freedom as self-mastery, equally liberated from the tyranny of nature and the tyranny of ideas. (In a footnote to Letter 19 Schiller acknowledges the possible misunderstandings caused by these two notions of freedom.)

Though the example set by the Greeks, Schiller submits, provides reason not to despair, he is well aware that experience and the historical record seem to speak volumes against the thesis. Still, they do so only if there is no transcendental path to a nonempirical, purely rational concept of beauty. Schiller accordingly proposes just such a path that takes its bearings from "the sheer potentialities" of human nature, potentialities that he juxtaposes with "what is absolute and unchanging" and the "necessary conditions" of human life. Though he feels no need to justify the considerable presuppositions built into this precarious move, what no doubt justifies it in his mind is

a fundamental analogy running throughout the *Letters*, namely, the analogousness of individual and political self-production to artistic production. In each case the reality in question can be conceived as the product of shaping something natural provided by experience, according to an idea that is, at least in regard to the initiative in question, irreducible to the respective experience of nature.

On the basis of this same analogy, the integrity of the reality (the production) in question demands that both nature and the idea—or, analogously, feeling and principle, the human condition and the human person—be given their due. Corresponding to this dual necessity are two basic laws of human nature, namely, “to externalize all that is within it, and give form to all that is outside it,” and two basic drives: a sensuous drive toward the material content of individual, momentary sensations, and a formal drive toward freedom in the form of universal, eternal laws. While the sensuous drive acts as a physical constraint and the formal drive as a moral constraint, the “task” of culture, Schiller submits, is to amplify each drive to the point where they have a moderating effect on one another. Departing from Kant and appropriating Johann Gottlieb Fichte’s accounts of a dialectical unity, Schiller declares that freedom requires, not the subordination of one drive to the other, but their coordination.

Schiller acknowledges the utopian character of the task. Still, he submits that there are moments in life when feeling and thinking merge, when human beings are able to realize both drives in a complementary way. These are the moments when human beings play. As Schiller famously puts it, “[M]an only plays when he is in the fullest sense of the word a human being, and he *isonly fully a human being when he plays*” (1993 [Letter 15, 1795] p. 131). (In Letter 27 Schiller gives a genealogy of play, from the physical play of an overflowing nature to the free play of human fantasy and association, culminating in aesthetic play with the capacity to transform sexual desire.) The play drive, as Schiller calls it, reconciles the otherwise competing sensuous and formal drives through its preoccupation with an object that combines their respective objects, life and form. In this way Schiller introduces his definition of beauty as a living form that is the object of the play drive. Precisely by yielding these moments of play, beauty is both a regenerative means to and a symbol of the consummate freedom that is, in his eyes, the destiny of humankind. Beauty here is not an empty (purposeless) form and the experience of it is not merely a matter of taste or the play of human faculties. Instead, it is a living form that embodies in a concrete, autonomous way the unity of feeling and principle, of

sense and reason. So conceived, beauty has a vitality that transcends human subjectivity without leaving it behind and yet, for this reason, holds an incomparable historical promise for humanity.

The already mentioned tension in Schiller’s conception of moral freedom takes on a new twist as Schiller describes freedom as the point where the sensuous and rational drives, far from being coordinated and facilitated, are said to be “canceled” (Letter 19). Further complicating matters, he gives an account of an “aesthetic condition” as a necessary means of predisposing human beings to a moral condition, “Man in his *physical* condition merely suffers the dominion of nature; he emancipates himself from this dominion in the *aesthetic* condition, and he acquires mastery over it in the *moral*” (1993 [Letter 24, 1795] p. 156). Still, if the aesthetic condition is now depicted as necessary for the transition to morality, its necessity is not something that one can leave behind. Beauty continues to be living proof “that a human being need not flee matter in order to manifest herself as spirit” (1993 [Letter 25, 1795], p. 165).

The transition from the aesthetic condition to the moral condition is supposedly far easier than the transition to the former from the physical condition. Hence, Schiller devotes his final remarks (Letters 26–28) to the role of “aesthetic semblance” in the former transition. Basic needs must be met, he notes, before aesthetic semblance can be indulged, though such indulgence is also a natural development of seeing and hearing. These two senses do not simply receive but help produce their objects. In the process, the play-drive develops, as people find enjoyment in mere semblance, as does the mimetic drive to shape and form this or that semblance into something relatively self-sufficient (though only relatively since it is a human product and subject to human dictates). As these drives develop, the realm of beauty expands but also gives further definition to the boundaries between semblance and reality. Moreover, only in this world of semblance does the artist enjoy sovereign rights. What makes the artist an artist and renders semblance aesthetic is a certain honesty (no pretense of being real) and autonomy (dispensing with all support from reality).

In the end, the aesthetic semblance is self-reflexive and self-redeeming. In an important respect art is the semblance of semblance, the illusion of illusion. The aesthetic education overturns a deficient, actual stage of human nature because art is capable of articulating ever higher human possibilities. Moreover, these are possibilities at the crossroads of the individual and the species. In



contrast to a strictly private sensual pleasure the enjoyment of semblance is a pleasurable activity that is inherently shareable, though not through some dictate of a *volonté générale*. Herein lies yet another side to the promise of beauty discussed earlier. Only in an aesthetic state (*Staat*) can we confront each other, not as enforcers of our respective rights (“the fearful kingdom of forces”) or as executors of our wills (“the sacred kingdom of laws”), but as free and equal citizens, “the third joyous kingdom of play and of semblance” (1993, [Letter 27, 1795], p. 176).

### POETS, PHILOSOPHY, AND PSYCHOLOGY

While Schiller concentrates in the *Letters* on art’s prospects of overcoming modernity’s alienating effects on humanity at large, his final major study, *On Naive and Sentimental Poetry* (1795–1796), turns to those effects on writers themselves. Naive poets, typified by ancient authors such as Homer, write effortlessly in a straightforward way without intruding themselves onto the scene, whereas “sentimental” (self-conscious) poets, so typical among modern writers like Ariosto, express their feelings about the scenes they depict. Characterizing the difference in terms of nature, Schiller explains, “The poet either is nature or will seek it. The former constitutes the ‘naive,’ the latter the ‘sentimental’ poet” (1993 [1795], p. 200). Thus, sentimental poets, in contrast to naive poets, are acutely aware of the difference between reality and their ideas and idealizations. Thus conflicted in their mode of feeling, they either mock reality in pathetic or playful satires, mourn the absence or loss of the ideal in elegies, or—most difficult of all—celebrate its future realization in idylls. Schiller’s use of the terms *naive* and *sentimental* is idiosyncratic; *naive* does not mean simplistic but direct, and *sentimental* does not mean maudlin but reflective.

Moreover, he construes the difference between these notions at times historically, at other times theoretically, to designate antithetical kinds of poetic consciousness in some contexts, and contrary traits within a single poet in others. For example, Goethe is a modern naive poet who is nonetheless capable of treating a theme “sentimentally,” as in his 1774 novel *Sorrows of the Young Werther*. (The contrast between naive and sentimental is in fact motivated, some argue, by Schiller’s attempt to come to terms with what he takes to be the difference between Goethe’s natural genius and his own more reflective, labored approach to writing.)

Nevertheless, in the first two parts of the essay, Schiller manages to accord each of these divergent literary

modes its due, while conceding “that neither the naive nor the sentimental character, considered in itself, can completely exhaust the ideal of beautiful humanity, an ideal that can only emerge from the intimate union of both” (1993 [1796], p. 249). That union itself is, Schiller adds, present only in “a few, rare individuals” since the difference between the naive and the sentimental poet is, he maintains, rooted in a broader difference as old as culture itself. Accordingly, in the third and final part of the essay, Schiller inscribes the difference between naive and sentimental poetry in a psychological profile of the difference between realists and idealists, that is, those who allow themselves to be determined in the end by nature or reason, respectively, be it in the form of the competing theoretical demands of common sense and speculation or the rival practical demands of happiness and nobility.

In “Concerning the Sublime” (first published in 1801 but begun around 1795) Schiller argues that sublimity must come to the aid of beauty in completing an aesthetic education, not least because nature’s intransigence defeats philosophy’s attempts to bring “what the moral world *demands* into harmony with what the real world *does*” (1993 [1801], p. 81). According to some critics, besides signaling a departure from the more optimistic (idealist) chord struck in the *Letters* and even the vestiges of a rationalist idea of harmony in *Naive and Sentimental Poetry*, this emphasis on philosophy’s limitations, with the grounding of realism and idealism in a “psychological antagonism,” explains why Schiller’s philosophical reflections on art largely come to a halt and he turns his attention once again to the stage. (One particularly noteworthy exception is the criticism of naturalism in the preface to the book version of the *Bride of Messina* in 1803, titled “On the Use of the Chorus in Tragedy,” an essay utilized by Friedrich Nietzsche in *The Birth of Tragedy* in 1871).

The influence of Schiller’s writings on German idealists and romantics is enormous. Shortly after the appearance of the *Letters*, Georg Wilhelm Friedrich Hegel writes Schelling that they are a “masterpiece,” and Johann Christian Friedrich Hölderlin makes plans to write his own “New Letters” on the same topic. Shaken in his neoclassicist beliefs by Schiller’s deft counterpoint of naive and sentimental poetry, Friedrich von Schlegel famously reconstrues them as “Classical” and “Romantic” poetry. Twenty years later, in his lectures on aesthetics, Hegel pays tribute to Schiller’s “great service of having broken through the Kantian subjectivity and abstraction and having dared to go beyond it, grasping unity and recon-

ciliation as the truth intellectually and realizing it artistically” (Hegel 1970 [1835], p. 89).

**See also** Aesthetics, History of; Beauty; Fichte, Johann Gottlieb; Goethe, Johann Wolfgang von; Hegel, Georg Wilhelm Friedrich; Hölderlin, Johann Christian Friedrich; Humboldt, Wilhelm von; Kant, Immanuel; Marx, Karl; Rousseau, Jean-Jacques; Schlegel, Friedrich von; Tragedy.

## Bibliography

### WORKS BY SCHILLER

*Schillers Werke*. 43 vols., edited by Julius Petersen et al. Weimar, Germany: Böhlau, 1943–.

*Naive and Sentimental Poetry, and On the Sublime: Two Essays*. Translated by Julius A. Elias. New York: Ungar, 1966.

*On the Aesthetic Education of Man*, edited and translated by Elizabeth M. Wilkinson and L. A. Willoughby. Oxford, U.K.: Clarendon Press, 1967.

Friedrich Schiller. “Über die ästhetische Erziehung des Menschen in einer Reihe von Briefen: Text, Materialien, Kommentar,” edited by Wolfgang Dörsing. Munich, Germany: Hanser, 1981.

*Essays*, edited by Walter Hinderer and Daniel O. Dahlstrom. New York: Continuum, 1993.

### WORKS ON SCHILLER

Hegel, Georg Wilhelm Friedrich. *Werke*. Vol. 13. Frankfurt am Main: Suhrkamp, 1970.

Martin, Nicholas. *Nietzsche and Schiller: Untimely Aesthetics*. Oxford, U.K.: Clarendon Press, 1996.

Reed, T. J. *Schiller*. New York: Oxford University Press, 1991.

Sharpe, Leslie. *Friedrich Schiller: Drama, Thought and Politics*. New York: Cambridge University Press, 1991.

Sharpe, Leslie. *Schiller's Aesthetic Essays: Two Centuries of Criticism*. Columbia, SC: Camden House, 1995.

*Daniel O. Dahlstrom (2005)*

## SCHLEGEL, FRIEDRICH VON

(1772–1829)

Friedrich von Schlegel, a critic and philosopher, whose writings spearheaded early German Romanticism, started out as a devotee of Greek poetry. Born to an illustrious literary family in Hanover and classically trained, Schlegel was an unhappy and unfocused student of law at Göttingen and Leipzig from 1790 to 1793, all the while piling up enormous gambling debts. Fleeing creditors and abandoning his legal studies, he moved in 1794 to Dresden where, inspired by Caroline Böhmer, his future sister-in-law, he launched his literary career with essays extolling

ancient poetry’s superiority to modern poetry. In “On the Study of Greek Poetry” (completed 1795, published 1797), he echoes Johann Joachim Winckelmann by attributing the greater unity, objectivity, and naturalness of ancient works to the Greeks’ single-minded pursuit of idealized beauty.

## PHILOSOPHY, CRITICISM, AND THE ROMANTIC TURN

Schlegel eventually wrote the *History of the Poetry of the Greeks and Romans*, but by the time the only volume was published in 1798, his view of modern poetry had changed. Already in his 1795 essay his admiration for William Shakespeare seems to belie his insistence on Sophocles’ superiority. His politics, too, though inspired by the ancients, were decidedly unconventional, as evidenced by his defense of the legitimacy of insurrection in his “Essay on the Concept of Republicanism” (1796), itself a critical review of Immanuel Kant’s “Toward Perpetual Peace” (1795). But it was chiefly Friedrich Schiller’s *On Naive and Sentimental Poetry* (1795–1796)—with its balanced judgment of the comparable virtues of ancient, “naive” and modern, “sentimental” (self-conscious) poetry and its reference to an even loftier poetry—that challenged Schlegel to reconsider his earlier views. Also like Schiller, Schlegel began to embrace Johann Gottlieb Fichte’s dialectical vindication of human dignity in the face of the threats posed to it by empiricism and mechanistic materialism. A growing awareness of Fichte’s impoverished view of nature eventually tempered this enthusiasm. Contrasting the “consistent empiricist” for whom everything sacred is “nonsense” with “mystics” as the real source of philosophy, Schlegel declares “Spinoza the best mystic known to us before Fichte” (*Kritische Ausgabe*, Vol. 18, p. 5). At the close of the eighteenth century Schlegel searched, much like Schelling and Georg Wilhelm Friedrich Hegel, for a philosophical path combining Benedict (Baruch) de Spinoza’s pantheistic naturalism with Fichte’s idealism.

Still, neither in Spinoza nor in Fichte did Schlegel find the sort of historical sensibility already exhibited in his early neoclassicist phase. This sensibility was accentuated in 1796 when, further signaling his departure from classicism, Schlegel begrudgingly accepted Kant’s argument that there are no objective rules for aesthetic judgments. Schlegel proposed that critics compensate for this lack of rules by being as comprehensively informed as possible of not only a writer’s but also an entire culture’s literary repertoire. At the same time he insisted that “criticism compares a work with its own ideal” (*Literary Note-*

books, p. 1135). This joint concern for a work's context and its sui generis character (exemplified by Schlegel's essays on Friedrich Heinrich Jacobi's *Woldemar* [1779], Georg Forster's works, and Gotthold Ephraim Lessing in 1797) would profoundly influence the development of hermeneutics by Wilhelm Dilthey and others.

In 1797 Schlegel moved to Berlin where close friendships with Friedrich Daniel Ernst Schleiermacher, Johann Ludwig Tieck, and Novalis (Friedrich Leopold von Hardenberg) gave rise to the new literary and philosophical movement eventually known as Romanticism. Its chief organ, the journal *Athenäum*, edited by Schlegel and his brother, August Wilhelm, contained Schlegel's most influential contributions to Romantic theory: "Fragments" and an essay on Johann Wolfgang von Goethe's *Wilhelm Meister* in 1798, and "Ideas" and "Dialogue on Poetry" in 1800. Another important source for Schlegel's theory is "Critical Fragments," printed in Johann Friedrich Reichardt's *Lyceum der schönen Künste* (1797). The form of fragments is itself a testament to the new theory's defiance of traditional literary theory. In memorable fashion Schlegel contrasts "Classical" with "Romantic" poetry, which disregards the traditional insistence on preserving purity of genre (epic, drama, and lyric). The novel (*Roman* in German) is, at least at first, paradigmatic for this theory that applauds the highly imaginative, genre-mixing fantasies (often with a love interest) typified by such Romance language writers as Dante Alighieri, Petrarch, and Miguel de Cervantes (but also by Shakespeare). In 1799 Schlegel provides his own example of a Romantic novel: *Lucinde*, a celebration of a complete but extramarital love, notoriously based on his affair with his future wife, the divorcée Dorothea Veit (Moses Mendelssohn's daughter).

## THE THEORY OF ROMANTIC POETRY

In *Athenäums-Fragment 116*, Schlegel's most influential account of Romantic poetry, he deems it "progressive universal poetry" because it aims not only to reunify all genres and connect poetry with philosophy but also to mingle and fuse "poetry and prose, genius and criticism, the poetry of the educated and the poetry of the people, to make poetry alive and social and to make life and society poetic, to poeticize wit, to fill and saturate the forms of art with matters of genuine cultural value" (*Athenäum I*, p. 220). To this end, a Romantic work is supposed to present sentimental but actual historical material in a witty, fantastic form ("an artfully ordered confusion") that is a synthesis of Eros and chaos, infinite unity and infinite fullness, mirroring nothing less than the universe

as a divine manifestation. The universe itself is conceived as a poem of the Godhead at this intersection of metaphysics and literary aesthetics.

In a good poem, as in reality, everything seems capricious and instinctive, though it is in fact necessary and deliberate. So, too, the Romantic artist must combine deadly seriousness with playfulness in a "constant self-parody," as Schlegel puts it. The model here is Socratic irony, a sense of the limitlessness of things and one's own limited capacity to express them, combined with the utter necessity of doing so.

In the final volume of *Athenäum* the emphasis on criticism and universality in the "Fragments" gives way to an enthusiasm for religion ("the all-animating world-soul of culture") and mythology (how "religion must appear in the world of language" (*Athenäum II*, p. 734, 740). In his "Ideas," which is deeply influenced by Schleiermacher, Schlegel touts the religious complementarity of poetry and philosophy that he also counterposes as realism and idealism, respectively. While claiming that "logic can develop into philosophy only through religion" and that "only someone who has his own religion can be an artist," Schlegel also insists paradoxically that "there is as yet no religion" (*Athenäum II*, p. 736, 751). Returning to this theme in "Dialogue on Poetry," he attributes the isolation of modern poets to their lack of a focal point such as ancient mythology provided ancient poets. He accordingly calls for the creation of a new mythology. This new mythology, like the ancient, would represent nature symbolically, though now against the background of philosophical idealism and the new physics (Schilling's philosophy of nature) and with an openness to the mythologies of the Orient.

## THE LATER WORKS

After 1800 Schlegel's fortunes initially took a turn for the worse. He failed as a lecturer on transcendental philosophy at the University of Jena and as a playwright, his collaboration with his brother ended with the publication of *Characterizations and Criticisms* (1801), and his relationships to other members of the Romantic movement deteriorated. Isolated and financially strapped, Schlegel moved in 1802 to Paris, where he published the periodical *Europa*, in which he influentially opposed classicism again, this time by championing the symbolism of early modern religious painters. Vainly looking for a professorship, Schlegel moved to Cologne in 1804, where he helped rediscover German Gothic architecture and published *On the Language and Wisdom of India* (1808). Seminal for the development of Sanskrit studies, comparative linguistics,

and Indian philosophy, this work also contained attacks on pantheism that introduced Schlegel's final, Catholic phase of thinking.

Following his conversion to Catholicism, Schlegel moved to Vienna and worked for the Austrian government (serving as Prince Klemens von Metternich's representative at the Diet of Frankfurt) but also found time to give well-received lectures: *On Modern History* (published 1811) and the monumental *History of Ancient and Modern Literature* (published 1815). From 1820 to 1823 he published the periodical *Concordia*, to which he contributed "Signature of the Age," a plea for an "organic" state headed by a strong monarchy and animated by "corporations," most prominently, the Church. In lectures on the philosophy of life, history, and language in his final years (published from 1828 to 1830), Schlegel challenged reigning philosophical systems—deduced, in his view, from merely a part of human consciousness—with a "Christian philosophy" grounded in the total, personal experience of a thinker as a believer.

**See also** Dante Alighieri; Dilthey, Wilhelm; Fichte, Johann Gottlieb; Goethe, Johann Wolfgang von; Hegel, Georg Wilhelm Friedrich; Jacobi, Friedrich Heinrich; Kant, Immanuel; Lessing, Gotthold Ephraim; Novalis; Petrarch; Romanticism; Schiller, Friedrich; Schleiermacher, Friedrich Daniel Ernst; Spinoza, Benedict (Baruch) de; Winckelmann, Johann Joachim.

### Bibliography

#### WORKS BY SCHLEGEL

- Kritische Friedrich Schlegel Ausgabe*. 35 vols., edited by Ernst Behler, et al. Munich: F. Schöningh, 1958–2002.
- Athenäum*, edited in two parts by Bernhard Sorg. Dortmund: Harenberg, 1989.
- Dialogue on Poetry and Literary Aphorisms*. Translated by Ernst Behler and Roman Struc. University Park: Pennsylvania State University Press, 1968.
- Literary Notebooks 1797–1801*, edited by Hans Eichner. London: Athlone, 1957.
- Lucinde and the Fragments*. Translated by Peter Firchow. Minneapolis: University of Minnesota Press, 1971. Partially reprinted in *Philosophical Fragments*. Translated by Peter Firchow. Minneapolis: University of Minnesota Press, 1991.
- "Essay on the Concept of Republicanism." In *Early Political Writings of the German Romantics*, edited and translated by Frederick C. Beiser. New York: Cambridge University Press, 1996.
- On the Study of Greek Poetry*. Translated by Stuart Barnett. Albany: SUNY Press, 2001.

#### WORKS ON SCHLEGEL

- Behler, Ernst. *German Romantic Literary Theory*. New York: Cambridge University Press, 1993.

Beiser, Frederick C. "Friedrich Schlegel's Absolute Idealism." In *German Idealism: The Struggle against Subjectivism, 1781–1801*, 435–461. Cambridge, MA: Harvard University Press, 2002.

Eichner, Hans. *Friedrich Schlegel*. New York: Twayne, 1970.

Frank, Manfred. *The Philosophical Foundations of Early German Romanticism*. Translated by Elizabeth Millán-Zaubert. Albany: SUNY Press, 2004.

*Daniel O. Dahlstrom (2005)*

## SCHLEIERMACHER, FRIEDRICH DANIEL ERNST (1768–1834)

Friedrich Daniel Ernst Schleiermacher was nineteenth-century Protestantism's great systematic theologian. It was he who marked the points of the compass for much of subsequent theology and philosophy of religion. Like St. Augustine, Schleiermacher desired to know God and the soul, and his place in the history of philosophy is due largely to the fact that he was able to state in modern language and concepts the great Augustinian conviction that religious faith is native to all human experience. Therefore, the knowledge of God and the knowledge of the soul are two orders of knowledge that must be distinguished but cannot be separated.

### LIFE

Schleiermacher was first and foremost a preacher and theologian, a church statesman, and an educator. He carried out his work as a philosopher in the context of the great idealist systems of Friedrich von Schelling, Johann Gottlieb Fichte, and G. W. F. Hegel, but instead of attempting to imitate these men he applied himself to the critical analysis of religion, both in its personal and societal manifestations, without reducing such experience to some form of philosophic intuition. The upbringing that his father, a Reformed clergyman, gave him and his early education in Moravian institutions set Schleiermacher upon this course. After studying at the university in Halle and taking his examinations for ordination in 1790, he served briefly as a private tutor to the family of Count Dohna in East Prussia and as a minister in the Prussian town of Landsberg. In 1796 Schleiermacher settled in Berlin as a preacher, became a close friend of Friedrich von Schlegel, and emerged as an interpreter of religion to the romantic worldview that Schlegel himself epitomized. *On Religion: Speeches to Its Cultured Despisers* (1799) gave Schleiermacher a national reputation at the age of thirty. The following year another publication, *Soliloquies*,

attested to Schleiermacher's thorough absorption of the spirit of romanticism, but at the same time it indicated the direction that his ethical interests were to take in the future, as in his *Grundlinien einer Kritik der bisherigen Sittenlehre* (Outline of a critique of previous ethical theory; 1803).

The relation between the religious and ethical dimensions of life constituted a major preoccupation of Schleiermacher's maturity, and it is here that his indebtedness to and divergence from Immanuel Kant are clearly evident. Of decisive importance during his Berlin sojourn was his embarking upon the translation of Plato, in the course of which his mind became imbued with the philosophy of the author of the *Republic*. By 1804 Schleiermacher was teaching philosophical ethics (philosophy of culture), theology, New Testament, and hermeneutics at Halle. By 1810 he was lecturing as professor of theology at the University of Berlin, where for the remainder of his life he taught dogmatic theology, New Testament theology and criticism, hermeneutics, practical theology, history of philosophy, ethics, and dialectics, to name only the more important of the wide variety of subjects with which he dealt. Concomitantly he held an appointment as preacher at the Dreifaltigkeitskirche, to which he attracted persons from all sections of Berlin, and from this pulpit he wielded a powerful moral influence on the nation. In ecclesiastical politics he labored for the union of the Lutheran and Reformed churches in Prussia, and in national politics he worked not only for stiffer resistance to French expansionism under Napoleon Bonaparte but for internal social reform.

*The Christian Faith (Der christliche Glaube nach den Grundsätzen der evangelischen Kirche im Zusammenhange dargestellt)* appeared in 1821–1822 and in revised form in 1830–1831. Together with the *Brief Outline of the Study of Theology* (1st edition, 1811) and the two open letters concerning the revised edition of *The Christian Faith* which Schleiermacher wrote to a close friend (*Sendschreiben über seine Glaubenslehre an Dr. Lücke*, 1829), *The Christian Faith* gives us not only Schleiermacher's thought on Christian doctrine and substantive theological issues but also his conception of the organization of the theological disciplines and of systematic theology itself. Schleiermacher made Protestant theology methodologically self-conscious.

## PHILOSOPHY OF CULTURE

Schleiermacher criticized Kant for tacitly making ethics into a "highest science" that ignored and devaluated the particular and idiosyncratic in human nature. Ethics,

Schleiermacher argued, is the discipline that has for its object "reason in history." Reason never appears except in historical personality—in the personalities of both individual persons and corporate persons. This position leads to a significant relaxation of the Kantian separation between practical reason, on the one hand, and the inclinations, temperament, talent, etc., on the other. Schleiermacher viewed these "accidents" and, indeed, the entire spatial, temporal embodiment of reason—apart from which we have no self-consciousness and hence no access to reason—not merely as the "place" of reason in its practical and theoretical functions but also as the organ of reason, by which reason itself is conditioned. The notion of a pure, universal reason could, therefore, be only a regulative concept for Schleiermacher.

Insofar as we consider reason in its practical capacity, as a willing or organizing activity, it is not the quest of virtue and autonomous assent to a self-imposed universal law that is foremost in view., but rather the sight of an ethical agent acting according to his own individuated rational nature. Moreover, the individuation of the ethical agent is accomplished not only by the "natural" accidents of time and place but also by the communities, societies, and institutions of which the individual person is the offspring. Schleiermacher presents the ethical agent as an end in himself, that is, as a good, who produces goods according to the peculiar law of his own unique nature. The doctrine of the highest good is formulated through the delineation of the relations of community and reciprocity in which such agents stand to each other, inheriting and endowing, receiving and bestowing. The primary forms in which these relations appear are the family, the nation, the church, the institutions of learning, and what Schleiermacher calls free sociality (*Geselligkeit*).

Nature and society affect reasoning in its theoretical as well as practical operations. When we think, we are conscious of engaging in an activity that is common to all men; nevertheless, our thinking, even at the most abstract level, as in thinking about thought itself, is in actuality predicated upon the specific organization of the physical means of sensation as well as upon the prior existence of a particular system of communication. The speculative activity of reason is thus conditioned by the natural medium in which it is individuated and shaped by the historical, moral character of the primary media (for example, a particular language) through which it maintains itself. Discourse is the means for the sociality of thinking, as Schleiermacher liked to say, and thinking is the inner side of speaking. He defined dialectic as the principles of correctly conducting a dialogue in the realm

of pure thinking and taught that all thinking proceeds in the form of dialogue or colloquy. On these grounds, Schleiermacher ruled out the possibility of an intuition of the absolute or of a highest science; the ideal and the real appear only as already informed by each other; pure spirit and matter lie outside of experience. Consequently, the ideal of a universal philosophy, for example, is nullified by the lack of a universal language and the impossibility of such.

The person, as the subject of the activities of thinking/knowing and of willing/doing, is more than a being composed of mind and body, individuated by time and space. A person not only is differentiated from others by nature and history but inwardly differentiates himself and acknowledges such an inward differentiation in all other human beings. That by virtue of which the person makes this inward differentiation is the *proprium* (*Eigen-thümlichkeit*). It is this property in each man that endows him with a life unity, an inalienable identity. Schleiermacher described this *proprium* as the peculiar organization that reason assumes for itself in each man. However, the life unity, or identity, of the individual person can never come to direct and full expression either in thinking/knowing or in willing/doing, although it accompanies and informs each of these rational activities. The self-consciousness that this sense of identity requires is a self-consciousness to be distinguished—though not isolated—from the forms of self-consciousness in which the subject is responding to or acting upon external objects.

Schleiermacher appropriates the word *feeling* for this form of self-consciousness, whose content is the given identity and unity of the self, incapable of being derived from others or surrendered to them. Feeling, thinking, and doing thus make up the three forms of consciousness that constitute the self-consciousness which distinguishes persons. Correspondingly, every person must be seen as a participant in the life of society in both his practical and theoretical functions, but he is also one whose *proprium* is wholly original. In a person whose feeling form of self-consciousness remains latent or inchoate, the sense of personal identity is deficient and personal consciousness is confused or immature. Such a person fails to contribute to the common or highest good; he is an inert reflection of his world, not one who moves and enriches it; he is a person in the formal sense but is destitute of spiritual life. Since, for Schleiermacher, religion is the most highly and fully developed mode of the feeling form of self-consciousness, all of human culture ultimately depends upon the cultivation of the religious life.

## RELIGION

In his earliest published work, the *Speeches*, Schleiermacher made ample use of the romantic preoccupation with the nature and value of individuality, but he qualified the world view of German romanticism in two important respects. First, an individual comes to self-knowledge only in the presence of other persons; hence the need to know and to express the self can be fulfilled only by observing and cultivating the morality of human community and communication. Second, the individual's cultivation of his own humanity—which the romantic accepted as a self-evident imperative—requires that he acknowledge his religious nature, as well as his aesthetic, scientific, and moral nature, and that he cultivate this side of his nature, or self-consciousness, by seeking out religious community. Schleiermacher's thesis, from 1799 to his death, was that man is a religious being. But since the individual must always appropriate his humanity in a fashion that is at once concordant with his generic identity and accordant with his own peculiar identity, religion is as much a problem for the individual as it is a natural endowment. In his mature thinking, as he came to align himself theologically with Augustine and John Calvin, Schleiermacher stressed not only the fact that man is a religious being but also the fact that the most fundamental, pervasive confusion inhibiting human consciousness is religious confusion. Thus, in his Christian theology, he described sin as the failure to maintain a clear distinction between that upon which men are entirely dependent, God, and that upon which men are only relatively dependent, namely, objects within the world.

In *The Christian Faith*, Schleiermacher stated that religion is a determination of feeling. More narrowly defined, it is a feeling of being absolutely dependent, and this feeling, he believed, is one and the same thing with consciousness of being in relation with God. A number of elements in this characterization need to be distinguished if Schleiermacher is to be understood. (1) The feeling of being absolutely dependent is also the feeling of identity, through which the individual is conscious of his inner uniqueness; in describing this feeling as one of being absolutely dependent, Schleiermacher was calling attention to the fact that the identity, or life unity, of the individual is an endowment which cannot be derived from any of the intellectual or volitional relations in which the self stands to other persons and forces, taken either singly or together. In this sense, the individual is utterly dependent, for the particular constitution of his existence, on a "whence" that cannot be rendered conceptually. Hence, the feeling of absolute dependence is not expressive of a

felt deficiency or of awe, as it is according to the interpretation of Rudolf Otto in *The Idea of the Holy*; nor is it wholly the same as Paul Tillich's conception of faith as being ultimately concerned about that which concerns us ultimately, since this concern is aroused in part by what Tillich called "nonbeing." (2) The feeling of being absolutely dependent—or "immediate self-consciousness" or "God-consciousness"; Schleiermacher regarded all three terms as equivalent—is discernible only because self-consciousness also involves thinking and willing, which are forms of rational relation between the person and his world, forms involving consciousness of "relative dependence" and "relative freedom." The feeling of being absolutely dependent is distinguishable from the feeling of relative dependence by virtue of the fact that in the latter a person stands in the relations of community and reciprocity with nature and society, while in the feeling of absolute dependence there is no reciprocity present. Consequently, there can be no consciousness of being in relation to God, apart from consciousness of being in relation to the world. (3) The original meaning of the word *God* is not a concept of perfect being, or the like, but the felt relation of absolute dependence. Hence, religion arises not in ideas, nor—for that matter—in willing, but in the immediate consciousness of what Schleiermacher described to Lücke as "an immediate existence-relationship." (4) In fact, then, religion is more than a determination of feeling; it is the name Schleiermacher gives to the personal self-consciousness in which the feeling of absolute dependence and consciousness of the world coexist and must achieve or receive a living, stable order.

The religion that Schleiermacher described in this way is a purely formal and abstract religion, which exists nowhere in actuality. In conformity with the principles we have outlined above, he insisted that religion always appears in a particular social and historical form. The great religions are religions bearing the stamp of their founders, and he defined Christianity as a monotheistic faith of the teleological variety in which everything is related to the redemption accomplished by Jesus of Nazareth. Everything in the outward, social, and institutional aspect of Christianity is related to its founder, and similarly, everything pertaining to the inner piety of the Christian is related to the historical figure of the redeemer. Thus, while Christianity is, without question, the religion on the basis of which Schleiermacher formed his understanding of all other religions, what is of more importance is that he was the first among modern theologians to perceive that Christianity is historical in two senses. Not only does it have a history, but each Christian becomes a Christian by appropriating to his total self-

consciousness the relation to Jesus Christ. Christ must become a part of the self-consciousness, or inner history, of the Christian. There is no part of the relation to God, Schleiermacher stated, in which the relation to Christ is not also actively present. Hence, Schleiermacher revived in his conception of the feeling of being absolutely dependent the Augustinian notion of the inseparability of the knowledge of the soul and the knowledge of God; at the same time he originated the distinctive form of modern Protestant theology—Christocentrism, or Christ as the center of the individual's inner religious consciousness.

**See also** Augustine, St.; Augustinianism; Calvin, John; Faith; Fichte, Johann Gottlieb; Hegel, Georg Wilhelm Friedrich; Kant, Immanuel; Otto, Rudolf; Philosophy of Religion, History of; Plato; Romanticism; Schelling, Friedrich Wilhelm Joseph von; Schlegel, Friedrich von; Tillich, Paul.

## Bibliography

### WORKS BY SCHLEIERMACHER

#### *Collected Works and Letters*

*Aus Schleiermachers Leben in Briefen*. 4 vols. Berlin, 1860–1863.

*Sämtliche Werke*. Berlin, 1835–1864. Published in three divisions: I, theological; II, sermons; III, philosophical and related subjects.

*Schleiermachers Briefwechsel mit J. Chr. Gass*. Berlin, 1852.

#### *English Translations*

*Brief Outline of the Study of Theology*. Translated by W. Farrer. Edinburgh, 1850.

*The Christian Faith*. Translated by H. R. Mackintosh and J. S. Stewart. Edinburgh, 1948; New York: Harper and Row, 1963.

*Christmas Eve: A Dialogue on the Celebration of Christmas*. Translated by W. Hastie. Edinburgh, 1890.

*The Life of Schleiermacher as Unfolded in His Autobiography and Letters*. 2 vols. Translated by F. Rowan. London, 1860.

"On the Discrepancy between the Sabellian and Athanasian Method of Representing the Doctrine of a Trinity in the Godhead." Translated by Moses Stuart. *Biblical Repository and Quarterly Observer* 6 (1835): 1–116.

*On Religion: Speeches to Its Cultured Despisers*. Translated by John Oman. London, 1893; New York, 1958.

*Selected Sermons of Friedrich Schleiermacher*. Translated by Mary F. Wilson. New York, n.d.

*Soliloquies*. Translated and edited by Horace Friess. Chicago, 1926.

#### *Critical Editions*

*Der christliche Glaube*, edited by Martin Redeker. 2 vols. Berlin, 1960. Contains appendices and Schleiermacher's own marginal notations.

*Grundriss der philosophischen Ethik*, edited by A. Twisten and F. M. Schiele. Leipzig, 1911.

- Hermeneutik, nach den Handschriften*, edited by Heinz Kimmerle. Heidelberg, 1959.
- Kurze Darstellung des theologischen Studiums*, edited by Heinrich Scholz. Hildesheim, 1961.
- Monologen*, edited by F. M. Schiele and Hermann Mulert. Leipzig, 1914.
- Reden über die Religion*, edited by G. C. B. Pünjer. Braunschweig, 1879.
- Schleiermachers Sendschreiben über seine Glaubenslehre an Lücke*, edited by Hermann Mulert. Giessen, 1908.
- Weihnachtsfeier*, edited by Hermann Mulert. Leipzig, 1908.

## WORKS ON SCHLEIERMACHER

- Barth, Karl. *From Rousseau to Ritschl*. London: SCM Press, 1959. A sophisticated chapter on Schleiermacher by his foremost theological critic.
- Brandt, Richard. *The Philosophy of Friedrich Schleiermacher*. New York: Harper & Brothers, 1941. The most detailed examination of Schleiermacher's philosophy in English.
- Dilthey, Wilhelm. *Leben Schleiermachers*. Berlin, 1870; 2nd ed., edited by H. Mulert. Berlin, 1922. The classic biographical work on Schleiermacher, but extends only to the period at Halle. Should be supplemented by Dilthey's articles in his *Gesammelte Schriften*, 2nd ed. Stuttgart, 1959–1960. Vols. IV and XII.
- Niebuhr, Richard R. *Schleiermacher on Christ and Religion*. New York: Scribners, 1964. An account of the central ideas in Schleiermacher's theology against the background of his philosophy.
- Süskind, Hermann. *Der Einfluss Schellings auf die Entwicklung von Schleiermachers System*. Tübingen: Mohr, 1909. Valuable account of the early development of Schleiermacher's thinking.
- Wehrung, Georg. *Die Dialektik Schleiermachers*. Tübingen: Mohr, 1920. A scholarly examination of the development of Schleiermacher's lectures on this subject.

*Richard R. Niebuhr (1967)*

## SCHLEIERMACHER, FRIEDRICH DANIEL ERNST [ADDENDUM]

In the past forty years there has been an explosion in research on Friedrich Daniel Ernst Schleiermacher in regard to both philosophical and theological dimensions of his thought. This entry is limited to discussing three issues of significance to philosophers: religious epistemology and the problem of religious pluralism, hermeneutics, and the question of the influence of Gottfried Wilhelm Leibniz, Immanuel Kant, and others on Schleiermacher's thought as a whole.

## RELIGIOUS EPISTEMOLOGY AND RELIGIOUS PLURALISM

Schleiermacher's contribution to the question of religious pluralism lies in his religious epistemology, which is developed in the first twenty-two chapters of *The Christian Faith* (1821–1822, second edition 1830/1999) as well as in *On Religion* (1799/1996; other editions followed in 1806, 1821, and 1831). In both, he offers a comprehensive theory of the nature of religion grounding it in experience. In *On Religion* he grounds religion in an original unity of consciousness that precedes the subject-object dichotomy, and in *The Christian Faith* the feeling of absolute dependence is grounded in immediate self-consciousness.

In *The Christian Faith* Schleiermacher explains that doctrines are expressions of this fundamental experience: Christian doctrines are “accounts of the Christian religious affections set forth in speech” (p. 76, § 15). This view has been labeled *experiential expressivism*. Christian doctrines are not a set of truth claims that are to be judged in virtue of their correspondence with reality, but are rather a human attempt to express in symbols the experience of absolute dependence. This original experience is immediate and is not itself conceptually structured, for any conceptual structure presupposes the subject-object dichotomy and thereby also one's counter-influence on that which is posited. God cannot be “given as an object exposed to our counter-influence, however slight this may be” (1999, p. 18; §4.4). As such, theological concepts and symbols are only indirect representations of one's consciousness of God. Given such an understanding of Christian doctrines, it is possible that two religions with differing symbols both adequately express the feeling of absolute dependence. Nevertheless, while the feeling of absolute dependence is not itself conceptually structured, it determines the way that one represents and knows oneself and the world around one. Hence, Schleiermacher states that “the world will be a different thing to a man according as he apprehends it from the standpoint of a God-consciousness completely paralyzed or of one absolutely paramount” (p. 267; §64.2).

## HERMENEUTICS

Schleiermacher's hermeneutics has also received a good deal of philosophical scrutiny. There are two diametrically opposed positions on the question of how it is possible to interpret a text or utterance. The first is the structuralist position: The meaning of any given utterance is determined by the publicly available meanings of the words that constitute it. Schleiermacher calls this the



“grammatical” element of language. According to the intentionalist position the meaning of an utterance lies in the intention of the speaker. The history and inner life of the speaker is of decisive importance in determining its significance. Schleiermacher’s hermeneutics, especially when understood in the context of his *Dialektik* (2001), offers a fruitful way to move beyond this impasse. According to Schleiermacher, one cannot strictly separate receptivity and spontaneity because both share a single underlying root. This plays a crucial role at several levels, the first being how one moves from sense-data to the ordinary world of tables and chairs. How the sense-data is organized will depend on the interpretive work of language: there is no bedrock given in receptivity. Ludwig Josef Johann Wittgenstein’s famous “duck-rabbit” is a useful example of this. Similarly, just as sense data provide no bedrock “given,” neither do the publicly available meanings of words. While language users begin from there, their own mental activity is important in shaping and sometimes even recasting those publicly available meanings. The level of the subject’s activity in shaping these meanings will vary from activity to activity, from high in aesthetic endeavors to low in scientific ones. For Schleiermacher, hermeneutics is “the art of understanding ... the ... discourse of another person correctly” (*Hermeneutics and Criticism*, p. 3). Both grammatical and psychological elements are vital to this task.

## RECEPTION OF THE PHILOSOPHICAL TRADITION

Lastly, a good deal of scholarship explores the systematic character of Schleiermacher’s thought and how it relates to preceding philosophical thought. In what ways was Schleiermacher influenced by the systems of Plato, Leibniz, Spinoza, and Kant? For instance, Schleiermacher’s *Dialektik* has received a good deal of scrutiny. Several scholars point to Schleiermacher’s Leibnizian heritage and its relation to Schleiermacher’s reception of foundational Kantian ideas. Specifically, Schleiermacher’s adoption of Leibniz’s complete concept, which contains all the predicates applicable to an individual, does not square with another idea essential to Schleiermacher’s system, namely that one is both a spontaneous and receptive being. Schleiermacher agreed with Kant that what is given to one through sensation is necessary, although not sufficient for knowledge. But if this is true, all true judgments cannot be analytic, as the Leibnizian tradition assumed. There is an important class of judgments that are synthetic: they are true in virtue of some third thing that one becomes aware of through one’s receptivity.

*See also* Religious Experience.

## Bibliography

### WORKS BY SCHLEIERMACHER

- The Christian Faith*. Translated by H. R. Mackintosh and J. S. Stewart. Edinburgh, 1948; New York, 1963.
- Kritische Gesamtausgabe*. 24 vols. to date. Edited by Hans-Joachim Birkner et al. Berlin: Walter de Gruyter, 1980–.
- On Religion: Speeches to Its Cultured Despisers* (1799). Translated with an introduction by Richard Crouter. New York: Cambridge University Press, 1996.
- Hermeneutics and Criticism and Other Writings*. Translated and edited by Andrew Bowie. New York: Cambridge University Press, 1998.
- Dialektik*, edited by Manfred Frank. Frankfurt am Main: Suhrkamp, 2001.
- Lectures on Philosophical Ethics*. Translated by Louise Adey Huih, edited by Robert B. Loudon. New York: Cambridge University Press, 2002.

### WORKS ON SCHLEIERMACHER

- Frank, Manfred. *Das Individuelle Allgemeine: Textstrukturierung und Interpretation Nach Schleiermacher*. 1 Aufl. Ed. Frankfurt am Main: Suhrkamp, 1989.
- Helmer, Christine, ed., in cooperation with Marjorie Suchoki, John Quiring, and Katie goetz. *Schleiermacher and Whitehead: Open Systems in Dialogue*. Berlin: Walter de Gruyter, 2004.
- Lamm, Julia. *The Living God: Schleiermacher’s Theological Appropriation of Spinoza*. University Park: Pennsylvania State University Press, 1996.
- Mariña, Jacqueline. *The Cambridge Companion to Friedrich Schleiermacher*. Cambridge, U.K.: Cambridge University Press, 2005.

*Jacqueline Mariña (2005)*

## SCHLICK, MORITZ (1882–1936)

Moritz Schlick, one of the founders of modern analytical philosophy and a guiding spirit of the Vienna circle of logical positivists, was born in Berlin. He was a direct descendant on his mother’s side of Ernst Moritz Arndt, the famous German patriot and political leader of the war of liberation against Napoleon Bonaparte. At the age of eighteen, Schlick entered the University of Berlin to study physics under Max Planck. He received his doctorate in 1904 with a dissertation on the reflection of light in a nonhomogeneous medium.

Schlick’s familiarity with the methods and criteria of research in the natural sciences left him dissatisfied with the epistemological notions both of neo-Kantianism, which then dominated the German universities, and of

Edmund Husserl's phenomenology, which had already become widely known. Instead, Schlick's starting point was the analyses carried out by Ernst Mach, Hermann von Helmholtz, and Henri Poincaré of the basic concepts and presuppositions of the individual sciences. His central interest at the time was the fundamental question of what is to be understood by knowledge.

From 1911 to 1917, Schlick served as lecturer and associate professor at the University of Rostock. In this period he published a series of works, among them his *Allgemeine Erkenntnislehre* (1918; 2nd ed., 1925). These works were devoted partly to a logically precise critical discussion of traditional philosophical conceptions and partly to an elaboration of new criteria for scientific knowledge which attracted considerable attention. In these publications Schlick already presented a first systematic account of his philosophical views.

In 1921 Schlick was named to a professorship at Kiel, and a year later he accepted a call to a chair in philosophy at the University of Vienna. These two years may thus be seen in retrospect as a kind of turning point in the history of philosophy. In 1921 Ludwig Wittgenstein had published his *Tractatus Logico-Philosophicus*, and in these same years the first writings of Rudolf Carnap appeared. Under the influence of Wittgenstein and Carnap, Schlick's philosophical views underwent a profound modification, which he later characterized by saying that he no longer saw the goal of philosophy as acquiring knowledge and presenting it as a system of propositions but, rather, as the application of a method. In applying its method, philosophy must take as its aim the discovery and understanding of the meaning of the statements, concepts, and formulations of problems of the special sciences, of philosophy, and of everyday life. When philosophy is understood in this manner, as Schlick emphasized in his French essay "L'école de Vienne et la philosophie traditionnelle" (*Travaux du IX<sup>ième</sup> Congrès International de Philosophie*, Paris, 1937), it resembles the method of Socrates, who constantly strove in his conversations to clarify the concepts, assertions, traditional notions, and ordinary modes of expression found in both the philosophy and the practical life of his time.

Schlick taught at the University of Vienna from 1922 until his death in 1936. During these years he twice made trips to the United States as a visiting professor. While in Vienna, Schlick published *Fragen der Ethik* (*The Problems of Ethics*, 1930), as well as numerous papers, most of which were later collected in various volumes. But his views were disseminated most effectively, perhaps, through the discussion society that he founded and that

acquired a worldwide reputation as the *Wiener Kreis*. Besides professional philosophers, regular participants in the meetings of the Vienna circle included primarily mathematicians and natural scientists but also psychologists and sociologists. They published a profusion of writings of their own, in which they applied the methods—constantly refined in discussion—of the new Vienna philosophy to the fundamental problems of scientific research.

Schlick was responsible for Carnap's appointment as lecturer at the University of Vienna. Another member of the Vienna circle was Kurt Gödel, who in this period published his famous proofs of the completeness of first-order logic and of the incompleteness of formal arithmetic. Numerous scholars from Germany, Poland, England, Norway, Sweden, and the United States visited the sessions of the Vienna circle and took part in its discussions. Conflicting views frequently were championed, but the application of the most rigorous logical tools to the positions under consideration was common to all the deliberations. These discussions thus turned out to be a genuine symposium in the classical sense of the term, and the international exchange of views that took place worked a transformation in the philosophical thought of the American and European universities.

On June 22, 1936, while on the way to his lecture in the main building of the University of Vienna, Schlick was fatally wounded by a deranged student. The motives for this act have never been fully clarified. The assailant had been under psychiatric observation for some time because of a previous attempt on Schlick's life. With the death of Schlick, the meetings of the Vienna circle came to a sudden end. The Austrian Ministry of Education, for its part, now embarked on a reactionary cultural policy that barred representatives of scientific, analytic philosophy from all official chairs in the universities. With few exceptions, the participants in the Vienna circle immigrated to England and America. The rigorous scientific requirements of the Vienna philosophy met with widespread sympathy in the West and in Poland and Scandinavia; as a result, philosophy as the "logic of knowledge" experienced a fruitful further development abroad.

In Austria, however, the philosophical movement initiated by Schlick encountered the uncompromising hostility of the state authorities. After the interruption caused by World War II, all the official chairs in the Austrian universities were systematically filled by speculative philosophers generally committed to a theological outlook. Only exceptionally was a representative of scientific philosophy able to qualify as a lecturer. But since lectur-

ers and associate or titular professors, unlike regular professors, are not paid a salary in Austria, the authorities had an effective economic means of compelling the unwanted logical analysts of knowledge to turn elsewhere. In practice, this resulted in a suppression of scientific philosophy that continues to exist to this very day. The necessary consequence of a policy so harmful to science has been a shocking decline in the level of scholarship. Psychologically, the only explanation for this reactionary course of isolating research from the rigorous demands of modern scientific philosophy is the fear that logico-mathematical or empirical scientific analysis might endanger some ideological position. In support of this view is the fact that the eastern European countries, which profess a diametrically opposed ideology, also keep Viennese logical positivism away from their chairs of learning out of the same medieval anxiety that prevails in Austria.

### CRITIQUE OF KANTIANISM

In his early work *Raum und Zeit in der gegenwärtigen Physik* (1917), Schlick presented a critical examination of the synthetic a priori character that Kantian transcendental philosophy attributed to propositions about space and time. Methodologically following the work of Poincaré and von Helmholtz, he based his thought primarily on the changes introduced by the theory of relativity into certain of the definitions and principles of classical physics. In conformity with scientific opinion of his time, Immanuel Kant had sought to establish the absolute validity of Newtonian mechanics by means of the theory of transcendental forms of intuition and of understanding. He regarded the presuppositions and basic principles of classical mechanics as necessary truths about empirical reality, that is, as synthetic a priori propositions. This conception had first been shaken by investigations of mathematicians. In consequence, doubt had also arisen regarding the synthetic a priori character of the general laws of physics. The theory of relativity made a final break with the synthetic a priori characterization of the foundations of Newtonian physics. According to relativity theory, statements about physical states (including propositions about physical space and physical time) are, as a consequence of the methods used by the natural sciences, empirical in character. That is, they are synthetic a posteriori propositions. Meanwhile, Poincaré had pointed to the possibility of interpreting general laws of nature, such as statements about physical space, as conventions or analytic propositions. Thus he had made evi-

dent the conventional nature of certain steps in the methodology of empirical research.

This systematic critique, confined at first to the foundations of mathematics and the natural sciences, was generalized by Schlick to all the basic problems of human knowledge. It thus became the basis of his philosophy in this initial period. In the *Allgemeine Erkenntnislehre* (1918), he made a critical study of all the propositions to which Kant and his followers had ascribed a synthetic a priori character. Schlick concluded that in all cases these propositions, where precisely formulated as logically necessary truths, are analytic in character; when, on the other hand, they are interpreted as statements with real content, they are empirical or synthetic a posteriori. There are no synthetic a priori propositions. Later, in his examination of foundational theories in logic and mathematics and of David Hilbert's formalism in particular, Schlick conceded that the possibility of synthetic a priori propositions in the realm of logico-mathematical forms must be left open. We are in no position to come to a final decision on this question. But even if necessarily valid propositions with content do exist—perhaps in the sense of the mathematical intuitionists—in the domain of logic and mathematics, they could never, Schlick stressed, be interpreted as absolutely valid statements about the empirically real world.

### CRITICAL REALISM

Schlick's view was that epistemology, in investigating the criteria of reality, is not obliged in the first instance to ask for absolutely true knowledge of reality. The Cartesian method of doubt leads merely to immediate data of experience, the establishment of which in no way suffices to answer the question "What is real?" Instead of seeking absolutely certain knowledge, we must address ourselves to the systems of propositions by the aid of which science seeks to describe reality, and through a critical examination expunge from these systems all propositions that are demonstrably false. The system that remains will then portray reality just as it is. Here, when we speak of the reality depicted by the natural sciences, we mean those phenomena described by true spatiotemporal propositions. Schlick identified the objects of empirical knowledge, thus characterized, with the Kantian thing-in-itself; he called his own philosophical position "critical realism."

According to Schlick, the method by which we arrive at knowledge of the spatiotemporally ordered world has the feature that whereas the truth of propositions about objective, empirical reality can in principle be established only hypothetically, the falsity of such propositions can in

some cases be demonstrated beyond question. It is interesting to note that Karl Popper's asymmetrical confirmation theory, which did not appear until some twenty years later, likewise attributes a kind of certainty to the disconfirmation of natural laws in contrast with the fact that full verification is unattainable.

In this first period of his philosophical development, Schlick regarded the controversy between idealism and realism as a factual issue which philosophical reflection could resolve. He believed that critical realism provided the correct answer, and he sought to substantiate this answer by a more precise characterization of what is to be understood by empirical knowledge. Knowledge is "knowledge of sameness." Something is cognized as something else, for example, a whale as a mammal. An especially important form of the knowledge of sameness is recognition. Memory outputs over short spans of time are a constitutive element of consciousness. Knowledge of sameness includes not only establishing the sameness or similarity of sense data, memory images, imagined ideas, and the like but also the rediscovery of certain conceptual orderings known, say, from mathematics in the relationships of empirical phenomena. Schlick did not consider the possibility that the study of empirical relationships might lead to the construction of new, hitherto unknown mathematical orders and that in such a case one might arrive at knowledge descriptive of reality that is not knowledge of sameness.

## LANGUAGE AND KNOWLEDGE

The problem of knowledge and its criteria had led Schlick to a further question: How is it possible to express knowledge linguistically? Scientific knowledge and insights, whether logico-mathematical or empirical, are presented in the form of sentences of some language. What conditions must be satisfied by these combinations of linguistic signs if they are to count as analytic or empirical sentences? In this earlier period Schlick's answer was the following: The languages employed in the sciences are designed to make possible the construction of unambiguous expressions that can be true or false. But this property of language presupposes the choice and establishment of rules according to which the linguistic signs are to be employed and to be strung out into expressions and sentences. If in using a language one does not heed the logical and linguistic rules set up for it, sign combinations will occur which, although they may appear on the surface to be sentences with a subject and a predicate, actually violate the rules for combining signs. Conse-

quently, they have no meaning and cannot be either true or false.

Applying this notion to philosophy, Schlick held that the theses of metaphysical systems are just such sequences of signs put together in a way that violates the logical rules of language. For this reason metaphysics is to be denied the status of scientific knowledge. But why does metaphysics disregard the logical rules of scientific languages in its linguistic formulations? Schlick thought the reason lay in the fact that whereas metaphysics endeavors to know reality, it does not seek to know the relations between the magnitudes characterizing states of affairs but strives to obtain knowledge of the content of phenomena. However, according to Schlick, only relations can be the object of knowledge—relations that reproduce the order of the phenomena and which include particulars on the number, sameness, similarity, and succession of the empirical data, as well as functional connections between measured quantities. The content of phenomena cannot be grasped by means of ordering relations, which are all that are at the disposal of the understanding. In Schlick's opinion, it is only through an intuitive, emotional experience that we can become acquainted with the actual content of reality. Metaphysics desires to know the "content" of real things, and it therefore finds itself compelled to use expressions from scientific languages in a manner contrary to the rules. For this reason the theses of metaphysics cannot have the character of meaningful propositions.

Schlick arrived at these views under the influence of the writings of Bertrand Russell and Hilbert, both of whom had by this time extensively treated the logical and linguistic foundations of mathematics. They clearly held that in mathematics questions about the logical and linguistic conditions for unambiguous statements must be put with special precision and exactness, but that these questions also affect the foundations of all scientific language systems and hence of scientific knowledge in general. Schlick was the first person to draw, on the basis of these insights into the foundations of logic and mathematics, consequences for epistemology as a whole and to undertake, by logical and linguistic means, the demarcation of a boundary between science and metaphysics.

## PHILOSOPHY AND REALITY

During his teaching career in Vienna, Schlick subjected the philosophical views he had published before 1922 to a fundamental reexamination. Influenced by Wittgenstein and Carnap, he no longer saw the task of philosophy as the acquisition of knowledge. Instead, philosophy,

through the application of logical analysis to the concepts, propositions, and methods of the separate sciences, should aim at reaching an understanding of knowledge as found in the individual disciplines and of its presuppositions. Schlick no longer treated realism and idealism as factually contradictory theses but, rather, as alternative ways of speaking; at most, one could ask which permits a simpler, more easily understood way of talking about the world of experience and about purely conceptual relationships. But if realism and idealism are interpreted as statements about something that exists, the realism-idealism antithesis becomes a “pseudo problem” to which neither a true nor a false answer can be given.

This conception was carried over by Schlick to certain problems in the foundations of physics. In his essay “Die Kausalität in der gegenwärtigen Physik” (1931, reprinted in *Gesammelte Aufsätze* and in *Gesetz, Kausalität und Wahrscheinlichkeit*), he cited the answer given by Werner Heisenberg when he was asked to what extent particles are real or unreal. Heisenberg had replied that whether or not one wished to label particles as really existing was simply a matter of taste (*Die physikalischen Prinzipien der Quantentheorie*, Leipzig, 1930, p. 15). In the systems of propositions that constitute physics, we speak only about the data of observation and the regularities they display, or we construct hypotheses and predictions about the occurrence of observable phenomena. Whether the terms *real* and *unreal* are applied to the observational data, to the hypothetical constituents, or to any other elements of the theories is, so far as the content of the system of propositions is concerned, of no consequence at all. Descriptions in terms of “real” and “unreal” can be omitted without any loss of asserted content. Whether one wishes to make use of these terms is merely a matter of convenience and simplicity in expression.

## PHILOSOPHICAL METHOD

Schlick generalized his analysis of modes of speech and ways of formulating questions into a philosophical method. Viewed from his new epistemological standpoint, numerous questions, especially in philosophy, turn out to be anchored in ordinary or scientific forms of speech, or in forms artificially created by metaphysics. The first step in Schlick’s method of analyzing knowledge consists in finding out the logical and linguistic rules governing the use of the expressions that occur in the problems, propositions, and forms of speech under study. Such a logical and syntactical critique may show that a certain expression, ordinarily assumed to have an unambiguous meaning, is being applied in accordance with dif-

ferent rules in different contexts and therefore is being used in different senses. A striking example is the concept of space. For a long time only one meaning was attributed to it, and the assumption was that the term *space* as employed in mathematics, physics, and psychology has the same meaning. The logical critique of language reveals that mathematical geometries represent analytic systems of relations, whereas physical space is described by means of a system of empirical laws that have as their content the order schema of possible positions and motions of physical bodies. Empirical sentences with different content describe the geometrical and metrical properties of psychological spaces—visual space, auditory space, tactile space, and the like. Similarly, in the case of such terms as *real*, *ideal*, *actual*, and *imaginary*, syntactical analysis yields different meanings corresponding to the different rules that govern the use of these expressions on various occasions. Failure to notice such differences of meaning often gives rise to philosophical problems which are then regarded as insoluble.

Thus the first step in the logical analysis of knowledge is to ascertain the rules for the linguistic use of the expressions under consideration. The second step is to study what meaning is to be ascribed to these expressions in a given complex of questions or system of propositions. Schlick called this the “interpretation” of the expressions, concepts, propositions, questions, or theories. If, for example, the first step in the analysis has shown that the word *real* is used in several senses, then the interpretation must determine which particular meaning the word has in, for instance, the sentence “Only that is real which is immediately experienced,” or in the sentence “The real is that which leaves traces behind,” or “The real is that which can be described by means of conjugate measured quantities.” The connection between the two steps in the method is manifest: The clarification of the possible meanings of an expression must precede the interpretation of it in a given context. According to Schlick, the understanding gained through interpretation is the insight for which philosophy strives.

Schlick applied his philosophical method, among other things, to the physical concepts of causality and energy and to the principles of causality and of the conservation of energy, which were still regarded as synthetic a priori propositions. Interpretation requires that in the case of “universally valid” sentences one must always ask whether one can conceive of conditions under which these sentences would have to be regarded as false. If they can be so regarded, then the empirical character of the sentences in question has been recognized. Schlick was

able to specify circumstances whose empirical confirmation is conceivable and under which both the principle of causality and the principle of the conservation of energy (as they are used within physics) would be termed invalid. Accordingly, he expressed the view—at a time when physicists were not yet of this opinion—that the two principles admitted of empirical testing. Subsequent research in physics has confirmed this view. At the same time, Schlick recognized that the concepts of causality and energy can also be defined in such a way that the principles of causality and of the conservation of energy become analytic sentences. It is this possibility that conventionalism exploits when it declares that general forms of laws are absolutely valid by convention. In a further application of his method, Schlick subjected Hans Driesch's vitalism and the general propositions both of psychology and of Husserl's phenomenology to an analytical critique. He arrived at the general conclusion that if the expressions these theories contain are precisely and properly clarified, the sentences in question take on either an analytic or an empirical character, but they never at one and the same time express synthetic and a priori propositions.

One criterion of meaning Schlick used in his analytical procedure was the criterion of verification that Schlick and others attributed to Wittgenstein. By this criterion, general laws of nature can have no significant content because they are not verifiable (or, as it is usually put, are not fully verifiable). This problem gave rise to wide-ranging discussions that went far beyond the Vienna circle. Essentially, Schlick supported Wittgenstein's view that natural laws are not themselves propositions but are to be understood as directives regarding the kind of sentences to be constructed in order to describe or predict individual cases of empirical phenomena. Directives cannot be true or false, so that on this interpretation the verification criterion is not applicable to the laws of nature. On several occasions Schlick characterized this interpretation of natural laws as not entirely satisfactory. But he did not find the opportunity for a definitive exposition of his own position.

## PRESUPPOSITIONS AND CONFIRMATION PROCEDURES

Schlick replied to certain criticisms of the philosophy of the Vienna circle. Doubt was expressed that the criteria of the analysis of knowledge are sufficient for distinguishing between analytic and empirical sentences or for drawing a boundary between metaphysics and the individual sciences. Extreme skeptics even questioned the possibility of

making such sharp distinctions at all. One argument used by critics concerned the presuppositions that are required whenever one attempts to specify the conditions for determining unambiguously the meaning of concepts and propositions or for deciding unambiguously the truth of analytic and empirical sentences. These presuppositions evade any formal characterization or any determination of their validity, and consequently they have a metaphysical character. Even if these ineluctable presuppositions are limited to the minimal performances of memory necessary for recognizing in a subsequent moment what meaning we have previously assigned to a given expression, the knowledge by recollection we thus presuppose is intuitive in kind and as impossible to check as the theses of metaphysics. Because of these problematic presuppositions, the logical positivist distinctions between analytic and empirical propositions and between scientific and metaphysical propositions cannot possess any validity.

Schlick analyzed these criticisms of recollections that cannot be checked but yet must be presupposed if consciousness, language, thought, and knowledge are to exist. The real problem of the logic of knowledge, he argued, consists in the fact that despite the inexact presuppositions of our methods of knowledge, we nevertheless do obtain exact scientific knowledge. It is wrong to conclude that because the recollections presupposed are unanalyzable and intuitive, the formal logico-mathematical derivations, concept formations, and principles or the empirical criteria of meaning and judgment are inaccurate. The exactness of scientific methods is anchored in proof procedures that guarantee an undeniable advance of knowledge in all the sciences. These procedures distinguish exact scientific knowledge from unverifiable metaphysical speculation. There are no such confirmation procedures for metaphysics, nor does it permit the application of scientific (logical or empirical) criteria of confirmation to its theses and methods. Consequently, in metaphysics there is no such thing as progress of knowledge. Thus the decisive criterion of exactness for the sciences is the advance in knowledge that can be gained through the process of testing, a criterion not satisfied by the speculative methods of metaphysics.

## ETHICS AND VALUE THEORY

Schlick also applied the method of the analysis of knowledge to problems of ethics and the theory of value. He concluded that the a priori arguments for absolute values do not fulfill the logical criteria of meaning. Only the value-ascribing forms of behavior actually found among

people, relative assignments of relative values, can be taken as the basis for ethical and other value systems. In Schlick's view, this sort of value analysis leads to a new kind of empirical foundation for eudaemonism. In his *Fragen der Ethik*, Schlick offered as the fundamental principle of an ethics so based the maxim "Increase your happiness" (*Mehre deine Glückseligkeit*).

Schlick's ethics has been widely criticized as superficial, on the ground that there can be morally objectionable happiness. To understand it correctly, one must take into account how he characterized the happiness that one should strive to increase. By happiness he meant the quiet, joyous assent that accompanies our actions when we carry out for its own sake some activity springing from our talents. This is the kind of activity that is to be evaluated as ethically worthwhile behavior. The joy in such activity resembles the joy of a child at play, and it should be regarded generally as the criterion for emotional and intellectual youthfulness. This youthfulness is not tied to physical age. Anyone who has found the activity proper to himself, and has thus experienced this quiet, joyous happiness, has realized the highest attainable ethical goal and will keep his youthfulness throughout his entire life. On this basis, Schlick rejected all varieties of ethical rigorism, including the Kantian system. No ethical worth can be attributed to actions undertaken from a mere sense of duty when such actions inspire only distaste and annoyance both beforehand and afterward. On the contrary, acting out of a sense of duty is ethically valuable only if a quiet satisfaction accompanies the action. Moral value, Schlick used to emphasize, attaches only to vital action; the sign of life is youthfulness, but we are young only when we act from joy. When the quiet, inner joyous assent accompanies our action, we fulfill the requirements of the highest principles of ethical value.

**See also** Critical Realism; Kantian Ethics; Logical Positivism; Neo-Kantianism.

## Bibliography

### WORKS BY SCHLICK

#### Books

- Raum und Zeit in der gegenwärtigen Physik. Zur Einführung in das Verständnis der Relativitäts- und Gravitationstheorie.* Berlin: Springer, 1917; 2nd ed., 1919. Translated by Henry L. Brose as *Space and Time in Contemporary Physics: An Introduction to the Theory of Relativity and Gravitation.* Oxford: Oxford University Press, 1920.
- Allgemeine Erkenntnistheorie.* Berlin: Springer, 1918; 2nd ed., 1925. Translated as *General Theory of Knowledge.* New York: Springer, 1974.

*Vom Sinn des Lebens.* Berlin, 1927.

*Fragen der Ethik.* Vienna: Springer, 1930. Translated by David Rynin as *Problems of Ethics.* New York: Prentice Hall, 1939.

*The Future of Philosophy.* Publications in Philosophy of the College of the Pacific, No. 1. Stockton, CA, 1932. Reprinted in *Gesammelte Aufsätze.*

*Gesammelte Aufsätze 1926–36* (Collected papers 1926–1936). Vienna: Gerold, 1938.

*Gesetz, Kausalität und Wahrscheinlichkeit* (Law, causality and probability). Vienna: Gerold, 1948.

*Grundzüge der Naturphilosophie*, edited by W. Hollitscher and J. Rauscher. Vienna: Gerold, 1948. Posthumous papers. Translated by Amethe Smeaton as *Philosophy of Nature.* New York: Philosophical Library, 1949.

*Aphorismen.* Vienna, 1962.

Schlick also edited, with Paul Hertz, the *Schriften zur Erkenntnistheorie* of Hermann von Helmholtz. Berlin: Springer, 1921.

#### Articles

- "Naturphilosophie." In *Lehrbuch der Philosophie*, edited by Max Dessoir, Vol. II, *Die Philosophie in ihren Einzelgebieten.* Berlin: Ullstein, 1925.
- "Gibt es ein materiales Apriori?" *Wissenschaftlicher Jahresbericht der philosophischen Gesellschaft an der Universität zu Wien für das Vereinsjahr 1930/31.* Translated by Wilfrid Sellars as "Is There a Factual A Priori?," in *Readings in Philosophical Analysis*, edited by Herbert Feigl and Wilfrid Sellars. New York: Appleton-Century-Crofts, 1949.
- "Die Wende der Philosophie." *Erkenntnis* 1 (1930–1931): 4–11. Reprinted in *Gesammelte Aufsätze.* Translated by David Rynin as "The Turning Point in Philosophy," in *Logical Positivism*, edited by A. J. Ayer. Glencoe, IL: Free Press, 1959.
- "Die Kausalität in der gegenwärtigen Physik." *Die Naturwissenschaften* 19 (1931). Reprinted in *Gesammelte Aufsätze* and *Gesetz, Kausalität und Wahrscheinlichkeit.* Translated by David Rynin as "Causality in Contemporary Physics." *British Journal for the Philosophy of Science* 12 (1961): 177–193 and 281–298.
- "Positivismus und Realismus." *Erkenntnis* 3 (1932–1933): 1–31. Reprinted in *Gesammelte Aufsätze.* Translated by David Rynin as "Positivism and Realism." In *Logical Positivism*, edited by A. J. Ayer. Glencoe, IL: Free Press, 1959.
- "Philosophie und Naturwissenschaft." *Erkenntnis* 4 (1934): 379–396.
- "Über das Fundament der Erkenntnis." *Erkenntnis* 4 (1934): 79–99. Reprinted in *Gesammelte Aufsätze*, Vienna: Gerald, 1935. Translated by David Rynin as "The Foundation of Knowledge," in *Logical Positivism*, edited by A. J. Ayer. Glencoe, IL: Free Press, 1959.
- "Facts and Propositions." *Analysis* 2 (1935). Reprinted in *Philosophy and Analysis*, edited by Margaret MacDonald. Oxford, 1954.
- "Unanswerable Questions?" *Philosopher* 13 (1935). Reprinted in *Gesammelte Aufsätze.* Vienna: Gerald, 1938.
- "Gesetz und Wahrscheinlichkeit." *Actes du Congrès International de Philosophie Scientifique, Paris 1935.* Actualités scientifiques et industrielles, no. 391, 46–57. Paris, 1936. Reprinted in *Gesammelte Aufsätze* (Vienna: Gerald,

1938), and in *Gesetz, Kausalität und Wahrscheinlichkeit* (Vienna: Gerald, 1948).

- “Sind die Naturgesetze Konventionen?” *Actes du Congrès International de Philosophie Scientifique, Paris 1935*. Actualités scientifiques et industrielles, no. 391, 8–17. Paris, 1936. Reprinted in *Gesammelte Aufsätze* (Vienna: Gerald, 1938), and in *Gesetz, Kausalität und Wahrscheinlichkeit* (Vienna: Gerald, 1948). Translated by Herbert Feigl and May Brodbeck as “Are Natural Laws Conventions?,” in *Readings in the Philosophy of Science*, edited by Herbert Feigl and May Brodbeck. New York: Appleton-Century-Crofts, 1953.
- “Meaning and Verification.” *Philosophical Review* 45 (1936). Reprinted in *Gesammelte Aufsätze* (Vienna: Gerald, 1938), and in Feigl and Sellars, *Readings in Philosophical Analysis* (New York: Appleton-Century-Crafts, 1949).
- “Quantentheorie und Erkennbarkeit der Natur.” *Erkenntnis* 6 (1936–1937): 317–326. Reprinted in *Gesammelte Aufsätze* (Vienna: Gerald, 1938), and in *Gesetz, Kausalität und Wahrscheinlichkeit* (Vienna: Gerald, 1948).
- “Form and Content: An Introduction to Philosophical Thinking. Three Lectures, Delivered in the University of London in Nov. 1932.” In *Gesammelte Aufsätze*.

#### WORKS ON SCHLICK

- Feigl, Herbert. “Moritz Schlick.” *Erkenntnis* 7 (1937–1939): 393–419.
- Popper, Karl. *The Logic of Scientific Discovery*. Translated by the author, with the assistance of Julius Freed and Lan Freed. London: Hutchinson, 1959.
- Reichenbach, Hans. “Moritz Schlick.” *Erkenntnis* 6 (1936): 141–142.
- Rynin, David. “Remarks on M. Schlick’s Essay ‘Positivism and Realism.’” *Synthese* 1 (1948–1949).

#### OTHER RECOMMENDED WORKS

- Gadol, Eugene T., ed. *Rationality and Science: A Memorial Volume for Moritz Schlick in Celebration of the Centennial of His Birth*. Wien; New York: Springer-Verlag, 1982.
- Haller, Rudolf. *Schlick und Neurath—Ein Symposium: Beiträge zum Internationalen philosophischen Symposium aus Anlass der 100. Wiederkehr der Geburtstage von Moritz Schlick (14.4.1882–22.6.1936) und Otto Neurath (10.12.1882 – 22.12.1945)*, Wien, 16.–20. Juni 1982. Amsterdam: Rodopi, 1982.
- Helmholtz, Hermann von. *Epistemological Writings: The Paul Hertz/Moritz Schlick Centenary Edition of 1921 with Notes and Commentary by the Editors*. Dordrecht, Holland; Boston: D. Reidel, 1977.
- Lewis, Joia A. *Positivism and Realism in the Writings of Moritz Schlick*. PhD diss., Indiana University, 1989.
- McGuinness, Brian. *Moritz Schlick*. Dordrecht; Boston: D. Reidel, 1985.
- McGuinness, Brian. *Zurück zu Schlick: Eine Neubewertung von Werk und Wirkung*. Wien: Holder-Pichler-Tempsky, 1985.
- Sarkar, Sahotra, ed. *Logical Empiricism at Its Peak: Schlick, Carnap, and Neurath*. New York: Garland, 1996.
- Schlick, Moritz. *Causality in Everyday Life and in Recent Science*. New York: Johnson Reprint, 1969.
- Schlick, Moritz. *General Theory of Knowledge*. Wien; New York: Springer-Verlag, 1974.

Schlick, Moritz. *Philosophical Papers*. Edited by Henk L. Mulder and Barbara F. B. van de Velde-Schlick. Dordrecht; Boston: D. Reidel, 1979.

Schlick, Moritz. *The Problems of Philosophy in Their Interconnection: Winter Semester Lectures, 1933–34*. Edited by Henk L. Mulder, Anne J. Kox, and Rainer Hegselmann. Dordrecht; Boston: D. Reidel, 1987.

Schlick, Moritz, and Bernd Philippi. *Philosophische Logik*. Frankfurt am Main: Suhrkamp, 1986.

Waismann, Friedrich. *Ethics and the Will: Essays*. Edited by Brian McGuinness and Joachim Schulte. Dordrecht; Boston: Kluwer Academic Publishers, 1994.

Waismann, Friedrich. *Wittgenstein und der Wiener Kreis*. Oxford: B. Blackwell, 1967.

**Béla Juhos (1967)**

**Translated by Albert E. Blumberg**

*Bibliography updated by Michael J. Farmer (2005)*

## SCHOLASTICISM

See *Augustinianism; Averroism; Medieval Philosophy; Occhamism; Scientia Media and Molinism; Scotism; Thomism*

## SCHOLZ, HEINRICH

(1884–1956)

Heinrich Scholz, the German theologian and logician, was born in Berlin. He professed an outspoken Platonism based on a profound knowledge of the history of metaphysics and of the logical works of Gottfried Wilhelm Leibniz, Bernard Bolzano, and Gottlob Frege. Scholz identified philosophy, in its original Platonic sense as the striving for universal knowledge, with the study of the foundations of mathematics and science. Thus, in *Was ist Philosophie?* (1940; *Mathesis Universalis*, pp. 341–387) he concluded, from Plato’s demand for knowledge of geometry and a mathematical astronomy, that the axiomatic method is required for universal knowledge. He regarded mathematical logic as developed by Leibniz, Bolzano, Frege, Bertrand Russell, and others as the “*epochale Gestalt*” of *metaphysica generalis*. He opposed formalism in logic because it failed to provide for the semantics of formal languages, and he opposed constructivism because of its arbitrary anthropocentric limitations of logic.

Scholz’s devotion to logic arose from a concern with metaphysics in theology. He studied theology at Berlin and philosophy at Erlangen, receiving a doctorate in philosophy from Erlangen with a dissertation on Friedrich Schleiermacher. He held the chair of systematic theology and philosophy of religion at Breslau from 1917 to 1919,



and then a chair of philosophy at Kiel. In his main systematic theological work, *Religionsphilosophie* (Berlin, 1921), he rejected subjective and existential foundations for religion. God is a transsubjective datum whose being is independent of any “leap of faith”; otherwise truth would be irrelevant to religion: “nothing remains but either to give up the solution to the problem of truth or to enter upon an entirely new course” (*Mathesis Universalis*, p. 13). By a “lucky accident,” the discovery of A. N. Whitehead and Russell’s *Principia Mathematica* in the library at Kiel, Scholz found his new course. From 1923 to 1928 he immersed himself in the study of logic, mathematics, and physics, and of their histories. His thoughts on metaphysics were galvanized, and he developed an enthusiasm for logical calculi rare even among mathematicians; it infused his later lectures and doubtless alienated those readers in Germany who were not quite convinced of the need to analyze Plato and other classical metaphysicians logically.

In 1929, his metamorphosis into a logician complete, Scholz assumed a chair of philosophy at Münster, which was transferred to the mathematical faculty in 1943 when he founded the Institut für mathematische Logik und Grundlagenforschung. This institute was inspired by the Warsaw school under Jan Łukasiewicz (whom Scholz later rescued from a Nazi concentration camp). But Scholz did not renounce theology. In “Das theologische Element im Beruf des logistischen Logikers” (1935; *Mathesis Universalis*, pp. 324–340) he likened his motives for undertaking *Grundlagenforschung* to the motives of an Augustinian theologian in search of illumination from the eternal forms. He undertook logical investigations of Anselm’s ontological argument (“Der Anselmische Gottesbeweis,” 1950; *Mathesis Universalis*, pp. 62–74) and of Augustine’s arithmetical proof (“Der Gottesgedanke in der Mathematik,” 1950; *Mathesis Universalis*, pp. 293–312).

Scholz wrote one of the first competent histories of logic, *Abriss der Geschichte der Logik* (Berlin, 1921; translated by Kurt F. Leidecker as *Concise History of Logic*, New York, 1961), based on the pioneering studies of Louis Couturat and Łukasiewicz. He exhibited what may be called a coincidence of logic and metaphysics through several works that together constitute in effect the first logically competent history of metaphysics. His “Logik, Grammatik, Metaphysik” (1944; *Mathesis Universalis*, pp. 399–438) discusses metaphysics in Aristotle, Leibniz, and Immanuel Kant. “Die mathematische Logik und die Metaphysik” (*Philosophisches Jahrbuch der Görres-Gesellschaft* 51 [1938]: 257–291), a 1938 lecture intended to convince a meeting of German Thomists of the impor-

tance of mathematical logic, discusses scholastic philosophy, Plato, and Aristotle. He discusses the fundamental importance of the axiomatic method for metaphysics in “Die Axiomatik der Alten” (1930; *Mathesis Universalis*, pp. 27–44), on Aristotle’s *Posterior Analytics*; in *Was ist Philosophie?*; and in *Die Wissenschaftslehre Bolzanos* (1937; *Mathesis Universalis*, pp. 219–267). Scholz regarded the *mathesis universalis* of René Descartes, Blaise Pascal, and Leibniz as of special importance in the history of metaphysics. He developed Leibniz’s metaphysical doctrines of identity and possibility in *Metaphysik als strenge Wissenschaft* (Cologne, 1941), a thorough treatment of the logic of identity, and in *Grundzüge der mathematischen Logik*, written in collaboration with Gisbert Hasenjaeger (Göttingen, 1961). In *Grundzüge*, logical truth is defined as that which is identical throughout all possible worlds. Scholz used this definition to explain the a priori (the pre-Kantian *Transzendente*): Possible (not necessarily actual) worlds constitute the logical frame for any description of the real world. Scholz’s “Einführung in die Kantische Philosophie,” a series of lectures given in 1943 and 1944 (*Mathesis Universalis*, pp. 152–218), was the first systematic treatment of Kant’s logical, mathematical, and physical doctrines to call upon both mathematical logic and physics. Of particular interest is Scholz’s account of how Kant came to reject the *mathesis universalis* because of Christian Wolff’s garbled presentation of Leibniz’s mathematical philosophy.

Scholz greatly admired the work of the Vienna circle, particularly that of Rudolf Carnap. However, he held that Platonism, especially in the form of classical mathematics, has been more useful to science than positivism, since it permits theoretical constructions more powerful than any offered by positivism. Positivism retards scientific growth. Thus, according to Scholz, modern relativity theory, even though positivistic tendencies helped lay its observational foundation, is Platonist because of its use of classical analysis. According to Scholz, the logic of Frege and Russell was adequate evidence that Platonism is feasible, and Alfred Tarski’s noneffective method of proof and his semantic definition of truth proved that Platonism can be given an absolutely rigorous foundation.

Scholz held that competence in metaphysics requires knowledge of mathematical logic, but he failed to convince most German metaphysicians. His works were ignored, and irrationalism exercised virtual hegemony in Germany during the Nazi era. (Even in the United States, his work was mentioned only in the *Journal of Symbolic Logic*.) Scholz saw language being employed as a poorly controlled, quasi-literary means of expression rather than

as a logical tool for grasping objective truth. He therefore engrossed himself in his technical work, the crowning achievement of which was the posthumously published *Grundzüge der mathematischen Logik*. This work deals extensively with the elements of logic; develops propositional logic, quantificational logic, and type-theoretical logic (this last is called “Russell-revised Platonism” because it functions as an ontological foundation for mathematics) in formalized syntactic and semantic metalanguages; and examines the questions of completeness and independence with respect to both effective and non-effective proof methods.

**See also** Anselm, St.; Aristotle; Augustine, St.; Bolzano, Bernard; Carnap, Rudolf; Couturat, Louis; Descartes, René; Frege, Gottlob; Kant, Immanuel; Leibniz, Gottfried Wilhelm; Łukasiewicz, Jan; Mathematics, Foundations of; Metaphysics, History of; Ontological Argument for the Existence of God; Pascal, Blaise; Plato; Platonism and the Platonic Tradition; Russell, Bertrand Arthur William; Schleiermacher, Friedrich Daniel Ernst; Tarski, Alfred; Thomism; Whitehead, Alfred North; Wolff, Christian.

### **Bibliography**

See *Mathesis Universalis*, edited by H. Hermes, F. Kambartel, and J. Ritter (Basel: B. Schwabe, 1961), a selective anthology with a comprehensive bibliography; *Glaube und Unglaube in der Weltgeschichte* (Leipzig: J. C. Hinrichs'sche Buchhandlung, 1911); *Die Religionsphilosophie des Als-ob* (Leipzig, 1921).

An article on Scholz, accompanied by a photograph, is Hans Hermes, “Heinrich Scholz zum 70. Geburtstag,” in *Mathematisch-Physikalische Semesterberichte* n.f. 5 (1955): 165–170.

*Eckehart Köhler (1967)*

## SCHOOL OF QOM, THE

The School of Qom refers to the tradition of theological institutions of Shi‘a learning in Qom, a city in southern Iran. Along with Meshhad in North Eastern Iran and Karbala and Najaf in Iraq, Qom is a major center of Shi‘ism, which houses the golden domed shrine of Fatimah, the holy site named for the sister of the eighth Imam who died in 816 in Qom and was buried there. Recently a few speculative theologians of this school proffered the theoretical foundation for a theocracy commonly labeled as an “Islamic Republic”; their views have become a cause célèbre in the Muslim world in a challenge of and a confrontation with the European cultural, economic, and

political dominance in many predominately Muslim states. In spite of its political charisma, Qom continues to be the source of research in the scholarship of the Shi‘a philosophical heritage and exports a number of both young and seasoned scholars to the most prestigious European academic centers.

### **BACKGROUND**

(a) Following the teachings of Nāṣir Khosrow (b. 1003–4) and Nāṣir al-Dīn al-Ṭūsī (1201–1274), traditional curricula of the school of Qom and its approach to Islamic studies integrate religious studies with philosophy and mysticism.

(b) In their ethics of self-realization and in their social philosophy, members of the school of Qom focus on philosophies of intentional processes and on analyses of mystical virtues (instead of “the golden mean”); they praise the use of archetypal memory (*dhikr*) and empathetic intimacy (*uns*) with the ultimate being. A key explicit view is the rejection of the Aristotelian depiction of time as an accident and the replacement of substance-event metaphysics with the so-called “process” ontology, expressed in Mullā Ṣadrā’s so-called theory of “substantial motion.”

(c) Members of the Qom school are actively engaged in an encounter with Europe; they do appreciate the development of European science as a continuation of Islamic sciences, and have mastered the art of application of computer technology to the humanities, such as the scanning of basic Shi‘a literature and internet communications.

### **MAJOR FIGURES AND CONTRIBUTIONS**

A majority of Iranian theologians were educated and taught in Qom. Salient doctrines of four major thinkers of this school follow:

**HUSSEIN TABATABA’I (1903–1981).** The most prominent thinker of this school is ‘Allameh Seyyed Muhammad Hussein Tabataba’i (hereafter “Tabataba’i”), a scholar of Shi‘a theology and teacher of recent major thinkers of the school of Qom. Although not directly involved in politics, his writings established the school’s theocratic agenda with key political implications.

A major aim of human actions is happiness, according to Tabataba’i—a desire that is only partially achievable in societal contexts that are regulated by laws that focus on external interrelations among human beings. Religious beliefs connecting the internal, intentional, and

spiritual bases of persons to the cosmos create an intimacy with the creator that complements the deficiencies of secular laws by providing total fulfillment of persons' needs, balancing tolerance with praxis. According to Tabataba'i, Islamic society goes beyond tolerance in recognizing the religious practices of other peoples of the book (Jews, Christians, and Zoroastrians) who live under the rule of an Islamic state and contribute to it through their taxes. However, the Muslim community is committed to jihad against those who knowingly reject the principle of unity, namely against rebellious Muslims, against enemies of the faith, and against those who transgress against Islam by occupying the Muslim homeland by force. Following Islamic traditional belief, Tabataba'i considers homicide a major sin against all humanity. For pragmatic purposes, truth may be hidden (the principle of *taqiyya*) when the expression of truth endangers the cause of religion. While speculation about religion is normally not recommended, Islam is open to new visions, inviting learned scholars to make innovations (*ijtihad*) in deducing philosophical points from archetypal monotheistic truths.

**MURTTAZA MUTAHHARI (1920–1979).** Mutahhari's major achievement was the dissemination of a clear, rational justification of the political views of the school of Qom to the Iranian Shi'a masses, as his books were printed by the tens of thousands and circulated as textbooks in many schools. Although Mutahhari is known primarily for his plan to refute communism, no less well known is his open acceptance of the advancement of European science and his caricature of the claim of supremacy of European secular philosophy. His research focuses on a number of politically important reforms. For example, he criticized the literal interpretation of sacred texts and advocated the rational adoption of religious archetypes to solve contemporary problems; he advocated the education of women; and he was receptive to the progress of science and technology. Significantly, he preferred political action as a specific innovative application of religious precepts carried out under the guidance of an exemplary political leader, and he eternalized the ethos of the martyrdom of Karbala, and in so doing providing an energizing rationale for the Islamic revolution.

**ROHALLAH KHOMEINI (1902–1989).** More than any member of this school, Khomeini permanently influenced the course of history of the Islamic world, claiming execution of basic Islamic principles such as Prophet Muhammad's agenda in the transformation of persons

from a tribal self to a member of a community of the faithful. In this tenor, a salient feature of Khomeini's political theory was his emphasis on how persons need to feel an intimate existential allegiance to Islam's spiritual nature in order to experience their religious societal self. The faithful are guided by the juridical authority in their participation in a revolution that has the following agenda: (a) to create a continuous confrontation with secular nationalism; (b) to challenge European secular capitalistic political and military imperialism that supports European puppet regimes in predominately Muslim countries; (c) to issue directives against the lives of those who transgress against Islam in any place in the world; and (d) to constitute—in the absence of an Imam (spirited leader)—a juridical authority that supercedes the authority of monarchs or even that of the elected president of a country.

#### PRESENT STATUS

Qom remains the major center of academic Shi'a research, where, in addition to Islamic studies, both male and female students study the works of philosophers such as Bertrand Russell, Ludwig Wittgenstein, and Martin Heidegger, and where mastery of computer technology applications to the humanities is expected of students.

*See also* Islamic Philosophy.

#### Bibliography

- Dabashi, Hamid. *Theology of Discontent: The Ideological Foundation of the Islamic Revolution in Iran*. New York: New York University Press, 1993.
- Khomeini, Rohollah. *Islam and Revolution*. Translated by Hamid Algar. Berkeley, CA: Mizan Press, 1981.
- Misbah Yazdi, Muhammad Taqi. *Philosophical Instructions: An Inquiry to Contemporary Islamic Philosophy*. Translated by M. Legenhausen and A. Sarvdalir. Binghamton, NY: Binghamton University, 1999.
- Mutahhari, Mortiza. *Fundamentals of Islamic Thought: God, Man, and Universe*. Translated by R. Campbell. Berkeley, CA: Mizan Press, 1985.
- Tabataba'i, 'Allameh Sayyed Mohamad Hosayn. *Islamic Teachings: An Overview*. Translated by R. Campbell. New York: Moustazafan Foundation, 2000.

*Parviz Morewedge (2005)*

#### SCHOPENHAUER, ARTHUR (1788–1860)

Arthur Schopenhauer was a German philosopher of pessimism who gave the will a leading place in his meta-

physics. He was born in Danzig. His father, a successful businessman of partly Dutch ancestry, was an admirer of Voltaire and was imbued with a keen dislike of absolutist governments. When Danzig surrendered to the Prussians in 1793, the family moved to Hamburg and remained there until the father's death (apparently by suicide) in 1805. Schopenhauer's mother was a novelist who in later years established a salon in Weimar, which brought him into contact with a number of literary figures, including Johann Wolfgang von Goethe. His relations with his mother, however, were bitter and antagonistic and eventually led to a more or less complete estrangement.

## EDUCATION

Schopenhauer's early education was somewhat unconventional. He spent two years in France in the charge of a friend of his father, and for another period he accompanied his parents on a prolonged tour of France, England (where he attended school in London for several months), Switzerland, and Austria. After his father's death he was tutored privately in the classics for a time and then entered the University of Göttingen as a medical student, studying, among other subjects, physics, chemistry, and botany. At Göttingen he first read Plato and Immanuel Kant, and the powerful and lasting impression their writings made upon him directed his interests decisively toward philosophy. In consequence he left Göttingen in 1811 for Berlin, which was at that time the chief philosophical center in Germany, and worked there for two years, attending the lectures of Johann Gottlieb Fichte and Friedrich Schleiermacher (both of whom he found profoundly disappointing) and making preparatory notes for a doctoral thesis. When the uprising against Napoleon Bonaparte led to the closing of the university, Schopenhauer, for whom nationalistic sentiment held little appeal, retired to Rudolstadt to write his thesis, subsequently published there in 1813 under the title of *Über die vierfache Wurzel des Satzes vom zureichenden Grunde* (*On the Fourfold Root of the Principle of Sufficient Reason*).

## EARLY CAREER

Apart from producing a short book on the perception of color, *Über das Sehn und die Farben* (Leipzig, 1816), which was inspired by a previous essay on the same subject by Goethe, Schopenhauer employed the next four years writing his principal work, *Die Welt als Wille und Vorstellung* (*The World as Will and Idea*). From the very first stages of the composition of this work, Schopenhauer believed that the ideas he was striving to express were of major importance, and when it was published at

Leipzig in 1818 (dated 1819), he was confident that its significance would immediately be recognized. In this expectation he was to be quickly disappointed; the scanty reviews his book received were generally tepid in tone, and the number of copies sold was small. Nevertheless, its publication helped him to obtain the post of lecturer at the University of Berlin, where he chose to give lectures at the same hours as G. W. F. Hegel, who was then at the height of his reputation and popularity. From the start, Schopenhauer advertised his opposition to Hegelian conceptions. He spoke of sophists who, having arisen after Kant, "first exhausted the thinking power of their time with barbarous and mysterious speech, then scared it away from philosophy and brought the study into discredit," and he made it clear that he regarded his own mission as one of repairing the damage that had been done. Schopenhauer's lectures, however, were a failure; Hegel's authority was too firmly established to be undermined in this manner, and Schopenhauer's audience dwindled away.

## LATER CAREER

Schopenhauer made no further attempt to establish himself academically. From then on he lived a solitary life, profoundly resentful at the lack of the recognition he felt to be his due and confirmed in his opinion that the dominant Hegelian philosophy was the product of a charlatan who, by an artful combination of sophistry and rhetoric, had succeeded in corrupting the intellects of an entire generation. Despite his disappointment, however, Schopenhauer continued to write, producing books that were in effect elaborations and developments of themes already adumbrated in his main work. He published an essay titled *Über den Willen in der Natur* (Frankfurt, 1836); and a volume on ethics and the problem of free will, *Die beiden Grundprobleme der Ethik* (Frankfurt, 1841), which contained the two essays "Über die Freiheit des Willens" (1839) and "Über die Grundlage der Moral" (1840). In 1844 he brought out a second edition of *Die Welt als Wille und Vorstellung*, greatly expanded by the addition of fifty supplementary chapters. He also contemplated translating Kant's *Critique of Pure Reason* into English and David Hume's *Dialogues concerning Natural Religion* (a work he greatly admired) into German. There can be little doubt that he would have performed both of these tasks well, for his knowledge of English was excellent; but unfortunately nothing came of either project. Finally, Schopenhauer published a collection of essays and aphorisms called *Parerga und Paralipomena* (2 vols., Berlin, 1851), and with this work he began to be widely

known. Discussions of his ideas appeared in foreign as well as in German periodicals, and his system was made the subject of lectures in a number of major European universities. By the time of his death in Frankfurt, he had a growing circle of admirers in England, Russia, and the United States, while nearer home the influence of his writings was soon to show itself in the work of such thinkers as Friedrich Nietzsche and Jakob Burckhardt.

## CHARACTER

Schopenhauer's personality, which is reflected in much of his writing, was complex and compounded of curiously diverse elements. Although intellectually self-assured to the point of arrogance, he had a brooding, introspective disposition, and he betrayed an extreme susceptibility to irrational fears and anxieties. Thus, he always slept with a loaded pistol near him, and he took compulsive precautions against disease; he once remarked that if nothing alarmed him, he grew alarmed at this very condition—"as if there must still be something of which I am only ignorant for a time." His manner could be truculent and overbearing; as many of his aphorisms make clear, his view of others was colored by a deep suspiciousness and cynicism, and his general outlook on life and existence was unrelievedly pessimistic. Yet this did not prevent him from taking pleasure in many things—art and music, good food and wine, travel, and, despite his notorious essay on the subject, women. And while he detested bores, in company that he found sympathetic he appears to have been a lively and entertaining talker, displaying a sharp, satirical wit.

## THE NATURE OF PHILOSOPHICAL THINKING

Schopenhauer's philosophy is best approached from a position that clearly recognizes his indebtedness to Kant, whom he believed to have been indisputably the greatest thinker of modern times. Schopenhauer's chief charge against his own philosophical contemporaries in Germany (Friedrich von Schelling, Fichte, and Hegel)—was that under the pretense of carrying forward and developing Kantian ideas, they had in fact attempted to philosophize in a fashion that Kant himself had ruled out as wholly inadmissible. For if Kant had shown anything, it was that metaphysical speculation in the old "transcendent" sense was useless as a means of achieving knowledge of what lay beyond all human experience. Such knowledge is in principle unattainable, and it followed that any philosopher, whatever his procedure might be, who tried to establish such things as the existence of God

and the immortality of the soul was engaged in a hopeless quest.

Rationalist metaphysicians like René Descartes had employed deductive a priori arguments in an endeavor to prove certain fundamental propositions of theology, and Kant had sufficiently exposed the inadequacy of these arguments by a series of devastating refutations. Yet according to Schopenhauer, Kant's strictures had not prevented some of his self-appointed successors from speaking as if they had mysterious access to truths necessarily outside the range of human cognition—a "little window opening on to the supernatural world," as it were. He suggested, too, that writing in this way appeared more expedient to many academic teachers of philosophy than the honest alternative of expounding truthfully and directly the antidogmatic theses contained in the *Critique of Pure Reason*.

While he accepted Kant's reasons for rejecting metaphysical theorizing in the sense described above, Schopenhauer was nevertheless far from wishing to claim that all philosophical speculation concerning the ultimate nature of the world must be deemed illicit and misconceived. The impulse to seek some general interpretation of reality and of the place of human existence within it was too deeply embedded in the human mind to be totally ignored or set aside. Man, Schopenhauer held, is an *animal metaphysicum*, a creature who cannot avoid wondering at the existence of the world and raising questions concerning its fundamental character and significance—questions that empirical science is unable adequately to resolve, for they lie beyond its sphere. Religion, it is true, attempts in its own way to meet this pervasive need, although not in a manner susceptible to rational justification or certification. For the tenets and concepts of religious faiths, whatever those who subscribe to them may believe to the contrary, can never be more than "allegories" or imaginative figures, and treating them as if they represented literal truths about a higher order of things leads straightway to manifest absurdities and contradictions.

By contrast, the concern of philosophical thinking is not with the metaphorical intimation of ideas that are beyond the grasp of the human intellect; rather, such thinking aims at truth *sensu proprio*. It follows, therefore, that any solution of "the riddle of the world" that philosophy purports to provide must not be one that involves overstepping the boundaries within which all human knowledge is set and confined. The determination of exactly where these boundaries lie is accordingly of pri-

mary importance as a preliminary to all philosophical inquiry.

## PERCEPTION AND THOUGHT

Schopenhauer's theory of knowledge may be said to start with Kant's distinction between *phenomena* (what appears to a perceiving mind) and *noumena* (things as they are in themselves). In our perceptual consciousness of the world, we are in fact aware of it only as mediated through our sense organs and intellect—a point Schopenhauer expressed by saying that, so conceived, the world is “idea” or “representation” (*Vorstellung*). Moreover, everything that presents itself to us in perception necessarily conforms to a certain formal and categorical framework that underlies and finds expression in all departments of our commonsense and scientific knowledge.

Thus Schopenhauer was at one with Kant in holding that the human mind cannot (as the British empiricists had suggested) be envisaged as a mere passive recipient of sense impressions, but on the contrary plays an essentially active part in shaping and organizing the sensory material. It is the structure of the intellect, comprising “sensitivity” and “understanding,” which ensures that this sensory material apprises us of a realm of external objective phenomena, spatially and temporally ordered and standing in determinate causal relations both with one another and with ourselves as percipients. Space and time as forms of sensibility, together with causality considered as the sole category of the understanding (here Schopenhauer diverged from Kant), are therefore “subjective in origin,” while at the same time they are necessary conditions of our knowledge of the world as idea. According to Schopenhauer, it is also the case that their valid employment is restricted to this sphere; they have no application to anything not given, or that could not be given, in sense experience.

Schopenhauer distinguished a further class of ideas, namely, what he termed “ideas of Reflection,” or sometimes “ideas of ideas” (*Vorstellungen von Vorstellungen*). It is in terms of these that we think about and communicate the contents of our phenomenal experience. In other words, they are the general concepts by virtue of which we can classify phenomena according to common features that are of interest or importance to us, forming thereby a conceptual structure or system that may be said to mirror or copy the empirical world. The function of this system is essentially a practical one; it provides a means of memorizing, and generalizing from, our observations of how things behave under varying conditions,

and hence of putting to use what we learn from experience.

Schopenhauer insisted, moreover, that this system cannot legitimately be separated from the foundation of empirical reality upon which it is based, and he claimed that concepts and abstract notions that cannot be traced back to experience are comparable to bank notes “issued by a firm which has nothing but other paper obligations to back it with.” Consequently, metaphysical theories that pretend to offer an account of the world purely a priori, and that in doing so employ terms or propositions not susceptible to empirical interpretation, are empty of cognitive content; they “move in the air without support.” Indeed, such theories often represent no more than the development, by laborious deductive steps, of the implications of a small group of initial axioms or definitions, yielding systems of empty tautologies.

Thus far, Schopenhauer would appear to have placed fairly stringent limits upon the scope of human inquiry. Attempts to transcend these limits by appealing to the resources of deductive reasoning alone are necessarily impossible, since they involve fundamentally wrong ideas concerning the nature of logical inference. These ideas can never provide us with information of which we were not previously cognizant, for such inference merely makes explicit what is already implicitly asserted in the premises from which it proceeds. Equally, there can be no justification for trying to extend the use of nonlogical, formative principles like the principle of causality in order to establish matters of nonempirical fact, after the manner of some earlier metaphysicians. Schopenhauer even accused Kant of inconsistency in this matter, on the ground that he wrote as though the existence of things-in-themselves, which for Kant are by definition incapable of being experienced, could be validly inferred from the phenomenal data, thereby disregarding his own prohibition. Nonetheless, Schopenhauer considered that the Kantian notion of the thing-in-itself remained a fertile one. Properly conceived, it offered the needed clue to the discovery of a legitimate and correct philosophical interpretation of existence.

## THE WILL

According to Schopenhauer, it is not true that the thing-in-itself, the noumenal reality that underlies the world of phenomenal appearances, is beyond the range of all possible human experience. To realize this, it is necessary to take account of the facts of self-consciousness, that is, our own intimate knowledge of ourselves. Self-awareness has two distinct aspects. From one point of view, namely, the

standpoint of ordinary perception, I cannot avoid regarding myself as an “object,” as much a physical entity as a building or a tree is. In this sense, I necessarily conform to the conditions that constitute the “world as idea” in general; I am a body that occupies space, endures through time, and causally responds to stimuli.

**INDIVIDUAL WILL.** My inner experience also assures me that I am nevertheless more than “an object among objects,” for I do not appear to myself under this aspect alone. I am also aware of myself from within as a self-moving, active being whose overt perceptible behavior directly expresses my will. This inner consciousness that each one of us has of himself as will is primitive and irreducible. Thus, Schopenhauer claimed that the will reveals itself immediately to everyone as the “in-itself” of his own phenomenal being and that the awareness we have of ourselves as will is quite different from the awareness we have of ourselves as body. At the same time, however, he emphatically denied that the operations of a man’s will and the movements he makes with his body are two distinct series of events—events of the first kind being thought of as causally productive of events of the second kind. Schopenhauer believed that dualistic conceptions of the relation of will and body, deriving largely from Descartes, had wrought havoc in philosophy, and he argued instead that a man’s body is simply the “objectification” of his will as it appears under the conditions of external perception; what I will and what in physical terms I do are one and the same thing, but viewed from different standpoints.

**THE WILL IN NATURE.** What has just been discussed represents the cornerstone of Schopenhauer’s metaphysic. For it was his contention that we should not assume the above distinction between the phenomenal appearance and the thing as it is in itself to apply only insofar as we ourselves are concerned. On the contrary, just as my own phenomenal being and activity is ultimately intelligible as the expression of my inner will, so may the rest of the phenomenal world be understood to share the same fundamental character that we recognize to be ours. Here was the “great extension” of the concept of will whereby Schopenhauer claimed that all phenomena—human and nonhuman, animate and inanimate—might be interpreted in a way that gave the world as a whole a new dimension of significance and that at the same time was not open to the insuperable objections vitiating traditional metaphysical doctrines.

The latter claim may reasonably be doubted. Schopenhauer often displayed considerable perspicacity

in detecting errors and inconsistencies in the theories of other philosophers, but he did not always show a comparable critical acumen with regard to his own ideas. Even so, the picture he drew of the world, in accordance with his conception of its inner essence, is not without a certain novelty and horrific fascination, standing as it does at the opposite pole from all those metaphysical systems that have, in one way or another, endeavored to present ultimate reality as if it were the incarnation of rational or moral order.

For Schopenhauer, the real was not the rational (as Hegel, for instance, implied that it was); on the contrary, “will” was for him the name of a nonrational force, a blind, striving power whose operations are without ultimate purpose or design. So portrayed, nature in all its aspects, ranging from the simplest physical structures to the most complex and highly developed organisms, takes on the character of an endless, and in the last analysis meaningless, struggle for existence, in which all is stress, conflict, and tension. The mechanistic models, the rationalistic schemes and constructions, in terms of which we find it useful to try to systematize the phenomenal data for scientific and practical purposes, merely serve to disguise from view the true nature of the underlying reality; the proper task of philosophy lies, not in seeking (as so many previous thinkers had sought) to reinforce these misconceptions by consoling and sophistical arguments, but rather in removing the veil of deception and setting the truth in a clear light.

**HUMAN NATURE.** As indicated above, Schopenhauer took as the starting point of his theory of the world the nature of man himself, regarded as the embodiment of will. Man is the microcosm in which all that is fundamental to reality as a whole (the macrocosm) may be plainly discerned. And it is in connection with what he wrote about human nature that Schopenhauer’s doctrine of the will can perhaps be most profitably considered. For this doctrine, far from being merely an extravagant philosophical fantasy, foreshadows much that was central to the later development of psychological theory; it represents a highly significant contribution with genuinely revolutionary implications.

*Will and intellect.* What Schopenhauer had to say on the subject of human nature revolved about his conception of the role of the intellect in human behavior. We like to suppose that in principle, everything we do lies within the province of our reason and is subject to our control; only if this is so can we deem ourselves to be truly our own masters. Traditionally, philosophers have given their

support to such beliefs; according to Schopenhauer, however, the situation is quite the reverse. For the will is not, as Descartes and others have taught, a sort of instrument or component of the intellectual faculty, mysteriously controlling our actions from on high by means of independent acts of rational choice. As has already been seen, Schopenhauer argued that will and body are simply the same thing viewed under different aspects, and he further claimed that the intellect, far from being the original source and spring of the will and the master of the body, is in fact no more than the will's servant and appendage. From an epistemological point of view, this governance of the intellect by the will manifests itself in the forms of knowledge under which the world appears to us—for example, as a causally governed system. To see things as causes or effects is to see them in terms of their potential uses, that is, as possible means to the gratification of the will.

**Motivation.** According to Schopenhauer, however, the primacy of will exhibits itself in a number of other important ways. Thus he gave various illustrations, drawn from everyday experience, of the manner in which we are often quite unaware of the true import and significance of our responses to circumstances and situations. Believing ourselves to be activated by some consideration that we find acceptable on moral or other grounds, we miss the real motive and might well be shocked or embarrassed if we knew it. Although we are inwardly and immediately aware of ourselves *as will*, our own consciously formulated conceptions of what we desire or what we are intending are, in fact, a highly unreliable guide when the question under consideration is *what we will*. Sometimes, indeed, Schopenhauer seems to have been making the extreme claim that conscious acts of choice never really determine behavior at all. He suggested in a number of instances that our conduct is not ultimately decided by resolves intellectually arrived at after weighing the pros and cons of alternative courses of action; the real decision is made by the will below the level of rationally reflective consciousness, the sole role of the intellect being to put before the will the various possibilities that lie open to the agent and to estimate the consequences that would ensue upon their actualization. In this sense, we never really form more than a “conjecture” of what we shall do in the future, although we often take such conjectures for resolves; what we have decided to do becomes finally clear to us only *a posteriori*, through the deed we perform. As it stands, this doctrine gives rise to obvious difficulties. Some cases doubtless occur that we should be inclined to describe in some such manner as Schopenhauer recommends, but it does not follow that every case of deliberate

action can be so characterized. Indeed, it may be claimed against all positions of this sort that it is only in virtue of our knowledge of what it is to act in accordance with consciously formed choices that the explanation of certain actions in terms of secret or concealed determinations of the will becomes intelligible.

**Unconscious mental activity.** The above-mentioned difficulties do not invalidate Schopenhauer's exceptionally perceptive and shrewd observations regarding much human motivation. These observations retain their importance even if the more bizarre speculations he based upon them are rejected; and Schopenhauer in fact connected them with a wider theory of human nature that, considering the time in which he wrote, manifested an astonishing prescience. According to this theory, the entire perspective in terms of which we are disposed to view our characters and doings is distorted. We customarily think of ourselves as being essentially free and rational agents, whereas in fact the principal sources and springs of our conduct consist in deep-lying tendencies and drives of whose character we are often wholly unaware. “Consciousness,” Schopenhauer wrote, “is the mere surface of our mind, of which, as of the earth, we do not know the inside but only the crust,” and in consequence we often put entirely false constructions upon the behavior in which these basic impulses are expressed. He suggested, moreover, that the ignorance we display, the rationalizations which in all innocence we provide, may themselves have a motive, although not one we are aware of. Thus, he frequently wrote of the will as preventing the rise to consciousness of thoughts and desires that, if known, would arouse feelings of humiliation, embarrassment, or shame. Another example of the same process is to be found in instances of memory failure. It is not a mere accident that we do not remember certain things, since there may be powerful inducements for us not to do so; events and experiences can be “completely suppressed,” becoming for us as if they had never taken place, simply because unconsciously we feel them to be unendurable. And in extreme cases this can lead to a form of insanity, with fantasies and delusions replacing what has thus been extruded from consciousness.

**Sexuality.** Sigmund Freud himself recognized the similarity between ideas like those above and some of the leading conceptions of psychoanalytical theory. Certainly there are striking parallels, and perhaps most obviously between what Schopenhauer had to say about the sexual instinct and the Freudian account of libido. For instance, Schopenhauer claimed that the sexual urge represents the “focus of the will.” Apart from the instinct to survive, it is



the most powerful motive of all and exercises a pervasive influence in every area of human life. Yet despite this, the amount of attention sexuality had received from most philosophers and psychologists had been remarkably small; it is as though a veil had been thrown over it, through which, however, the subject kept showing through. Nevertheless, Schopenhauer was far from extolling the operations of the sexual drive. Although he thought it necessary to expose honestly the stark reality that human beings seek to hide by falsely romanticizing and idealizing their primitive passions, he also made it clear that he considered sexuality to be a source of great mischief and suffering. Thus he referred to it as a “demon” that “strives to pervert, confuse and overthrow everything,” and spoke of sexual desires as being inherently incapable of achieving lasting satisfaction; according to Schopenhauer, the end of love is always disillusion. In other words, here, as elsewhere, conformity to the dictates of the will ultimately results in unhappiness, which is the universal condition of human existence.

**PESSIMISM AND ANTIRATIONALISM.** In sum, Schopenhauer’s doctrine of the will constituted, in a variety of ways, a reaction against the then dominant eighteenth-century, or “Enlightenment,” conceptions of human nature. He not only rejected the Cartesian belief in the primacy of intellect or reason in man, but also, by implication, repudiated the “mechanistic” model according to which writers like Hume sought to explain human personality and motivation in terms of the combination and association of atomistically conceived impressions and ideas. In place of this model, he substituted one of dynamic drive and function that was oriented toward the biological rather than the physical sciences and that stressed the importance of unconscious rather than conscious mental processes. Furthermore, Schopenhauer’s writings represent a complete departure from the strain of optimism that underlay so much eighteenth-century thinking about history and society. Schopenhauer utterly rejected such ideas as the inevitability of human progress and the perfectibility of man and replaced them with a picture of humankind in general as doomed to an eternal round of torment and misery. Radical changes in the social structure, however “scientifically” applied, would solve nothing, for the evil condition of life as we find it is merely the reflection of the aggressive and libidinous urges rooted in our own natures. All that can usefully be employed are certain palliatives in the form of social and legal controls that give the individual minimal protection against the incursions of his neighbors; and with such measures men have long been familiar.

## ART AND AESTHETIC EXPERIENCE

The preeminent position that Schopenhauer assigned to art (certainly no other major philosopher has elevated it to a higher status) is not difficult to understand in the light of his general theory. In this theory, our modes of knowledge and understanding, as well as the activities in which we normally engage, are regarded as being determined by the will. Scientific inquiry was the supreme instance of this, since (Schopenhauer believed) its essential function was one of providing, through the discovery of empirical uniformities, practical techniques for satisfying our wants and desires.

**THE AESTHETIC ATTITUDE.** The artist’s concern, however, is not with action, or the possibility of action, at all, but with what Schopenhauer termed “contemplation” or “will-less perception.” This type of perception must not be confused with perception of the ordinary everyday kind, wherein things are looked at from the standpoint of practical interest and appear under the aspect of particular phenomenal objects. For it is the mark of aesthetic contemplation that in the enjoyment of artistic experience “we keep the sabbath of the penal servitude of willing”; the world is seen in abstraction from the various aims, desires, and anxieties that accompany our normal apprehension of it, with the result that it presents itself to us in a completely different light.

It is a further consequence of such detachment (and on this point Schopenhauer followed Kant) that all judgments of taste or aesthetic value are disinterested: They cannot have as their basis some titillation of sensual appetite, for instance; nor can they be grounded upon considerations of social utility, or even of moral purpose. To speak of a natural scene or of a work of art or literature as “beautiful” is to judge it in and for itself, and quite outside the framework of cause and consequence within which our ordinary perceptual judgments have their natural place and from which they derive their significance.

**THE AESTHETIC OBJECT.** The claim that aesthetic awareness presupposes a distinctive attitude of mind and attention is clearly separable from the contention, also advanced by Schopenhauer, that in such awareness the content of our experience is of a radically different kind from that involved in ordinary sense perception. Surprising as it may seem in the light of some of his earlier pronouncements, Schopenhauer held that the subjective conditions that define and universally determine our perception at the everyday level are wholly in abeyance in the case of aesthetic apprehension, and that to this complete

“change in the subject” there is a corresponding change in the object. As aesthetic observers, we are no longer confronted with a multiplicity of individual things and events that are spatiotemporally and causally interrelated, but instead are presented with the “permanent essential forms of the world and all its phenomena,” which Schopenhauer termed the “Ideas” (*Ideen*). This conception of fundamental Ideas, which Schopenhauer adapted from Plato to serve the purposes of his own, very different, theory of art, helps us to understand why he regarded art not merely as a kind of knowledge, but as a kind of knowledge vastly superior to any found in the sphere of the natural sciences. In his view, the natural sciences can never do more than discover regularities at the stage of phenomenal appearance, whereas works of genuine art exhibit to the beholder the nature of the archetypal forms of which the particular phenomena of sense perception are necessarily incomplete and inadequate expressions. Artistic productions may, in fact, be said to be the vehicles through which the artist communicates his profound discoveries and insights and thereby enables others to share his vision.

The notion that the proper objects of artistic perception are Platonic Ideas in the sense described above gives rise to obvious objections. It certainly fits somewhat uneasily into Schopenhauer’s system insofar as that originally seemed to be based upon the postulate that phenomenal representation and noumenal will between them exhaust the field of possible human knowledge. And quite apart from this, the theory of Ideas raises problems on its own account. It appears paradoxical, for instance, to suggest that a picture of, say, apples in a bowl is not a picture of things of the sort we can all see and touch in the ordinary way, but of something set mysteriously apart from these and situated in a realm beyond the range of normal vision.

Even so, it is at least to Schopenhauer’s credit that he recognized some of the difficulties presented by much that we are prone to think and say about artistic portrayals of experience. The concept of perception, for instance, seems to play a significantly different role in the context of aesthetic appraisal and criticism from the role it plays in other contexts. Again, the specific sense in which certain art forms (painting, for example) are concerned with “representing” reality is notoriously difficult to analyze. The claim that the artist sees something literally distinct from what we ordinarily see is, no doubt, hard to defend; on the other hand, the (different) claim that he sees and is able to portray ordinary things in unfamiliar ways, and under fresh and revealing aspects, appears to contain an obvious truth.

Schopenhauer himself never clearly distinguished between these two claims. Theoretically he subscribed to the first, but much that he said in his discussion of concrete cases accords better with the second. Not only did he often stress the particularity of the artist’s observation of phenomena; he also suggested that the artist’s unique mode of presenting individual objects, scenes, or situations succeeds in illuminating for us whole ranges of our experience to which we have previously been blind. He argued, however, that it would be a mistake to suppose that we can ever convey by verbal description what we learn from our direct acquaintance with particular works of art. For what these works communicate will in the end always elude anything we try to say about them. “The transition from the Idea to the concept,” he wrote, “is always a fall.”

**MUSIC.** Schopenhauer thought that all forms of artistic activity—with one important exception—could be understood and explained in terms of his theory of Ideas. The exception was music. Music is not concerned with the representation of phenomena or the fundamental forms that underlie phenomena, but has as its subject the will itself, the nature of which it expresses directly and immediately. Thus, of all the arts, music stands closest to the ultimate reality of things that we all bear within ourselves and speaks “the universal imageless language of the heart.” Schopenhauer’s ideas, in this instance and in general, produced a deep impression upon Richard Wagner, who in his opera *Tristan und Isolde* tried to realize in musical form the leading conceptions of Schopenhauer’s theory of the world. It is a curious irony that Schopenhauer, far from reciprocating Wagner’s admiration, spoke of his music with actual distaste.

## ETHICS AND MYSTICISM

Although the world, viewed from a purely contemplative standpoint, presents a spectacle that can be aesthetically enjoyed, it does not follow that the operations of the agency which underlies all that we perceive can afford us any kind of moral guidance or solace. On the contrary, the ethical significance of existence lies in its ultimate horror. Unlike many other metaphysicians, Schopenhauer concluded from his system, not that we should gratefully seek to make our lives conform to the pattern implicit in the nature of reality, but rather that true salvation consists in a total rejection of this pattern. The moral worth of individuals lies in their capacity to liberate themselves from the pressures and urges of the rapacious will.

**INALTERABILITY OF CHARACTER.** It is not altogether easy to see how liberation is possible. Schopenhauer had claimed that human beings, like everything else in nature, are in essence expressions of will. How, then, can they become otherwise? Furthermore, he insisted upon a strictly deterministic interpretation of human character and action, one that makes the type of freedom of choice postulated by traditional libertarian doctrines inconceivable. What a person does is always and necessarily a manifestation of his inner disposition, which remains fixed and is unalterable by any resolutions he may form to be different. The individual discovers what he is really like by observing his behavior over the course of his life. He will find that this behavior conforms to certain invariant patterns of reaction and response, so that if the same circumstances recur, his conduct in the face of them will be the same as it was before. Such consistent behavior patterns are the outward manifestation of the individual noumenal essence, or timeless character, which each man is in himself—a conception Schopenhauer claimed to have derived from Kant's discussion of the foundations of moral responsibility, though the consequences he drew from it were in fact far removed from any drawn by Kant. Nor can some of these consequences be said to have been logically very happy; for instance, Schopenhauer seems to have employed the notion of a man's character so elastically that it ruled out the possibility of any imaginable state of affairs falsifying his thesis concerning its innate and unchangeable nature.

**ETHICAL VARIATION.** Schopenhauer's claim that a man cannot change his character at will does not, however, commit him to the view that the dispositions of different individuals do not show significant ethical variations. For an explanation of the fact that there are good as well as evil persons in the world, he returned to the fundamental tenets of his metaphysic. It is a feature of the good, as contrasted with the self-centered or egotistical, individual that he comprehends himself and his relations with others from a "higher" standpoint, which enables him to recognize, however obscurely or inarticulately, the common unitary nature shared by all things. Egoism rests upon the assumption that the individual is a self-sufficient unit, to which all else is foreign. But the individual appears to be set apart from his fellows by an impassable gulf only when apprehended in accordance with the spatiotemporal scheme that informs our everyday "will-governed" way of looking at things.

A profounder insight, such as is exhibited intuitively in the behavior of the just and compassionate man who "draws less distinction between himself and others than is

usually done," involves awareness of the illusory character of the phenomenal world. Those who possess this awareness no longer see their fellow creatures as alien objects to be overcome or manipulated in pursuit of their own ego-centric aims, but rather as "themselves once more," homogeneous with their own being and nature. Thus, in the last analysis, the distinction between virtue and vice has its source in radically different modes of viewing those around us; and this distinction could, Schopenhauer believed, be adequately explicated and justified in the terms provided by his own philosophical system.

**DENIAL OF THE WILL.** Schopenhauer frequently quoted the Brahman formula, *tat tvam asi* ("that thou art"), when discussing the metaphysical unity of things that underlies the realm of appearance. Indeed, all his writings on ethical and related subjects show affinities with the doctrines advanced in the Upaniṣads and in Buddhist texts—affinities that he freely acknowledged. Like the Indian teachers, he considered all human life to be enmeshed in suffering, and following them, he often used the word *māyā* to refer to the illusory phenomenal world to which, as empirical individuals, we belong. Total release from the enslavement of the will, as compared with the identification of himself with others that is displayed in the conduct of the morally good man, in fact occurs only when a person finally ceases to feel any attachment to earthly things and when all desire to participate in the life of the world completely vanishes. Such an attitude of mind, which Schopenhauer attributed to ascetics and mystics of all times, becomes possible when a man's will "turns and denies itself," and when what in the eyes of ordinary men is the very essence and substance of reality appears to him as "nothing."

But Schopenhauer was insistent that this "turning of the will," which is a highly mysterious process, is not something a man can bring about through his own deliberate volition, since the process involves the complete "abolition" of his previous personality. This "turning of the will" comes to him, as it were, "from outside" and springs from an insight that wholly transcends the will and the world. Such mystical insight, moreover, is necessarily incommunicable and indescribable; all knowledge, including that attainable by philosophy, here reaches its limit, and we are left with only "myths and meaningless words" that express no positive content. "The nature of things before or beyond the world, and consequently beyond the will," Schopenhauer declared at the close of his main work, "is open to no investigation." The end of philosophy is silence.

## IMPORTANCE AND INFLUENCE

Schopenhauer's critics have not failed to draw attention to discrepancies and inconsistencies in his system. These certainly exist, and his natural clarity of expression, which contrasts so sharply with the obscure and cloudy terminology favored by his philosophical contemporaries in Germany, makes them comparatively easy to detect. On the other hand, these discrepancies should not be allowed to stand in the way of a proper appreciation of what was important and influential in Schopenhauer's thought. The nineteenth century witnessed a decline in the fascination that achievements in physics and mathematics had previously exercised over philosophy, and there was a tendency in speculative thought to explore new ways of interpreting and conceptualizing human life and experience. In this development Schopenhauer played a central role. Both through his theory of will, with its psychological implications, and also through the new metaphysical status he gave to art, he helped to bring about a profound shift in the intellectual and imaginative climate.

In this connection, the impression made by his ideas upon novelists such as Lev Tolstoy, Joseph Conrad, Marcel Proust, and Thomas Mann is particularly noteworthy. Among philosophers, the impact of Schopenhauer's thought was weaker and certainly never approached that produced by Hegel's writings; while in more recent times, when philosophical speculation in general has been at a discount, he has attracted little interest. Yet such neglect is undeserved, and the significance of his contribution should not be underestimated. He realized more fully than the majority of his contemporaries the implications of the Kantian critique of traditional metaphysics, and some of the things he himself had to say about the nature of a priori knowledge have a strikingly modern ring. Again, it is worth emphasizing his "instrumentalist" view of human thinking, which anticipated William James and the American pragmatist school, and also his highly perceptive attacks upon the Cartesian theory of personality and self-consciousness, which in important respects foreshadowed present-day approaches to problems in the philosophy of mind. (In particular, his theory of the double knowledge we have of ourselves as agents in the world has interesting contemporary analogues.)

Finally, it should be remembered that possibly the greatest philosopher of modern times, Ludwig Wittgenstein, read Schopenhauer and was influenced by him. The extent of this influence appears most clearly in the notebooks Wittgenstein kept during World War I (*Notebooks 1914–1916*, translated by G. E. M. Anscombe, Oxford,

1961), but signs of it are also to be found in the *Tractatus Logico-Philosophicus* (translated by D. F. Pears and B. F. McGuinness, London, 1961), particularly in the sections on ethics and the limits of language in the latter part of the work.

**See also** Kant, Immanuel; Pessimism and Optimism.

## Bibliography

### WORKS BY SCHOPENHAUER

#### German Editions

*Sämtliche Werke*, edited by Arthur Hübscher. 7 vols.

Wiesbaden: F. A. Brockhaus, 1972; 4th edition, 1988.

*Der handschriftliche Nachlaß*, edited by Arthur Hübscher. 5 vols. Frankfurt-on-Main: Kramer, 1970.

*Philosophische Vorlesungen. Aus dem handschriftlichen Nachlaß*, edited by Volker Spierling. 4 vols. Munich and Zurich: Piper, 1984–1986.

#### English Translations

*Schopenhauer's Early Fourfold Root*. Translated by F. C. White. Aldershot: Avebury, 1997.

*On Vision and Colors*. Translated by Eric F. J. Payne. Oxford: Berg, 1994.

*The World as Will and Representation*. Translated by Eric F. J. Payne. 2 vols. New York: Dover, 1969.

*On the Will in Nature*, edited by David E. Cartwright.

Translated by Eric F. J. Payne. New York and Oxford: Berg, 1992.

*Prize Essay on the Freedom of the Will*, edited by Günter Zöllner. Translated by Eric F. J. Payne. Cambridge, U.K.: Cambridge University Press, 1999.

*On the Basis of Morality*. Translated by Eric F. J. Payne. Providence and Oxford: Berghahn, 1995.

*On the Fourfold Root of the Principle of Sufficient Reason*.

Translated by Eric F. J. Payne. LaSalle, IL: Open Court, 1974.

*Parerga and Paralipomena*. Translated by Eric F. J. Payne. 2 vols. New York: Dover, 1966.

*Manuscript Remains*. Translated by Eric F. J. Payne. 4 vols. Oxford: Berg, 1988.

### WORKS ON SCHOPENHAUER

Atwell, John E. *Schopenhauer on the Character of the World: The Metaphysics of Will*. Berkeley: University of California Press, 1995.

Atwell, John E. *Schopenhauer: The Human Character*. Philadelphia: Temple University Press, 1990.

Copleston, Frederick. *Arthur Schopenhauer: Philosopher of Pessimism*. London: Barnes and Noble, 1975.

Fox, Michael, ed. *Schopenhauer: His Philosophical Achievement*. Brighton: Harvester Press, 1980.

Gardiner, Patrick. *Schopenhauer*. Middlesex: Penguin Books, 1967; repr. Bristol: Thoemmes Press, 1997.

Hamlyn, D. W. *Schopenhauer*. London: Routledge & Kegan Paul, 1980.

Hübscher, Arthur. *The Philosophy of Arthur Schopenhauer in Its Intellectual Context: Thinker Against the Tide*. Translated by Joachim T. Baer and David E. Cartwright. Lewiston, NY: Edwin Mellen Press, 1989.

- Janaway, Christopher, ed. *The Cambridge Companion to Schopenhauer*. Cambridge, U.K.: Cambridge University Press, 1999.
- Janaway, Christopher. *Schopenhauer*. Oxford: Oxford University Press, 1994.
- Janaway, Christopher. *Self and World in Schopenhauer's Philosophy*. Oxford: Clarendon Press, 1989.
- Magee, Bryan. *The Philosophy of Schopenhauer*. Oxford: Clarendon Press, 1983; 2nd ed., 1997.
- Malter, Rudolf. *Arthur Schopenhauer. Transzendentalphilosophie und Metaphysik des Willens*. Stuttgart and Bad Cannstatt: Frommann-Holzboog, 1991.
- Malter, Rudolf. *Der Eine Gedanke. Hinführung zur Philosophie Arthur Schopenhauers*. Darmstadt: Wissenschaftliche Buchgesellschaft, 1988.
- Neymeyr, Barbara. *Ästhetische Autonomie als Abnormität: Kritische Analysen zu Schopenhauers Ästhetik im Horizont seiner Willensmetaphysik*. Berlin: Walter de Gruyter, 1996.
- Safranski, Rüdiger. *Schopenhauer and the Wild Years of Philosophy*. Translated by Ewald Osers. London: Weidenfeld and Nicolson, 1989.
- Simmel, Georg. *Schopenhauer and Nietzsche*. Translated by Helmut Loiskandl, Deena Weinstein, and Michael Weinstein. Amherst: University of Massachusetts Press, 1986.
- Spielring, Volker, ed. *Materialien zu Schopenhauers "Die Welt als Wille und Vorstellung"*. Frankfurt-on-Main: Suhrkamp, 1984.
- von der Luft, Eric, ed. *Schopenhauer: New Essays in Honor of His 200th Birthday*. Lewiston, NY: Edwin Mellen Press, 1988.
- White, F. C. *On Schopenhauer's Fourfold Root of the Principle of Sufficient Reason*. Leiden: E. J. Brill, 1992.
- Young, Julian. *Willing and Unwilling: A Study in the Philosophy of Arthur Schopenhauer*. Dordrecht: Martinus Nijhoff, 1987.

**Patrick Gardiner (1967)**

*Bibliography by Günter Zöller (2005)*

## SCHRÖDINGER, ERWIN

(1887–1961)

Erwin Schrödinger was born in Vienna, Austria. After his years at the Gymnasium, where he was given a strong education in classics and in science, he studied physics and mathematics at the university of Vienna from 1906. His major teachers were the successors of Ludwig Boltzmann: Franz Exner and Fritz Hasenöhr. Schrödinger's early interest for philosophy is evident in several manuscripts of this period, which contain reflections about Greek and Indian thought and British empiricism. He was then awarded the D. Phil. Degree in 1910 and became assistant experimental physicist in Exner's laboratory in 1911. From this date until 1922, he worked on several subjects, including atmospheric radioactivity, statistical physics, psycho-physics of sensations, general relativity, and atomic physics.

At the end of World War I, in which Schrödinger served as an artillery officer in the Austrian army, he devoted one year to studying philosophy. He wrote down his philosophical reflections later, during the summer of 1925, in an essay that became part one of his book *My View of the World*. After brief appointments in various German universities, he became full professor of theoretical physics at the university of Zurich in 1922. In the autumn of 1925, he formulated wave mechanics, construed as a development and alteration of Louis de Broglie's ideas. In 1926 Schrödinger published classic papers in which he formulated and solved the "Schrödinger equation" and demonstrated the empirical equivalence between wave mechanics and Werner Heisenberg's matrix mechanics.

Schrödinger then succeeded Max Planck at the prestigious chair of theoretical physics of Berlin in 1927. But he left Berlin to go to Oxford in mid-1933, a few months after Hitler's rise to power. The same year, he shared the Nobel Prize for physics with Paul Dirac. During his two years at Oxford, he wrote several important papers about the interpretation of quantum mechanics, presenting for the first time the concept of "entanglement" of states and the "cat paradox." Schrödinger then accepted an appointment at the University of Graz in Austria in 1936. But with the advent of the "Anschluss" in 1938, Schrödinger had to flee once more from the Nazis. After spending some time in Vatican and in Belgium, he received an appointment in 1940 at the Dublin Institute for Advanced Studies. The Institute was founded mainly for him by Eamon de Valera, then prime minister of the Irish Republic. During his stay in Ireland, Schrödinger devoted his work to unified field theories (in Einstein's spirit), to renewed reflections on the interpretation of quantum mechanics, and also to conferences for a broader audience. His well-known books *What Is Life?*, *Nature and the Greeks*, and *Mind and Matter* arose from these conferences. In 1956, Schrödinger returned to Austria, where he retired in 1958. He died in Vienna in January 1961.

### PHILOSOPHY OF PHYSICS

The key to Schrödinger's philosophy of physics (especially quantum mechanics) is contained in a letter to Arthur Eddington of March 22, 1940. There, Schrödinger insists that Ernst Mach's radical empiricism and Ludwig Boltzmann's taste for rational "pictures" are not mutually exclusive strategies. He regarded Mach's empiricism as a good guide to *tabula rasa* whenever unwarranted old intellectual constructs hinder a proper understanding of new physical phenomena. But this is only the first step of

research. Boltzmann's urge to picture must be the second step. Indeed, "forming absolutely clear, almost naively clear and detailed "pictures" allows one "to be quite sure of avoiding contradictory assumptions." (*The Interpretation of Quantum Mechanics*, 1995, p. 121).

Schrödinger used both methods. He was clearly inspired by Mach's method when he criticized vehemently the old-fashioned concept of "particle" construed as a small permanent material body. He formulated his criticism as early as 1913, when he first heard of Bohr's model of the atom, and then refined it throughout his career. According to Schrödinger, the concept of an object is constructed out of actual observations complemented with appropriately selected virtual observations. But if the interpolation of arbitrarily numerous virtual observations is not allowed by the most advanced predictive theory, then the very process of construction collapses, and the corresponding object *cannot be said to exist*. For elementary particles, "Observations are to be regarded as discrete, disconnected events. Between them there are gaps which we cannot fill in" (*Science and Humanism*, p. 27). We cannot fill them in according to a trajectory pattern, because of Heisenberg's uncertainty relations. But if there is no trajectory, the discrete events cannot be tied up into a spatio-temporal continuant. Therefore, the idea that these scattered events reveal some permanent being is a sort of kinetic *illusion*: particles *do not exist*.

After this Machian preliminary move, however, Schrödinger activated the Boltzmannian side of his philosophy of physics. To him, without a precise picture, scientific thought is threatened with ambiguity. Yet the picture must not be taken as mere mimicry of "things out there." It is nothing more than the most efficient *mental tool* we have, with no ontological implications. This is the status Schrödinger ascribed to his *wave function* in the 1950s, after having apparently held a naively realist belief in the existence of  $\psi$ -waves in 1926. His mature view of wave functions was expressed in *Science and Humanism* (p. 40): "We do give a complete description, continuous in space and time ... a description of something. But we do not claim that this 'something' is the observed or observable facts; and still less do we claim that we thus describe what nature ... really is." Yet the description, or picture, must be taken seriously in view of its epistemological value. Its continuous evolution according to the Schrödinger equation and the entanglement between wave functions must be allowed to develop throughout without any sudden "reduction of the state." The only constraint to be exerted on this picture is that it must have some connection with experimentally observable

events. But to secure this connection, it is sufficient to use either a rule about expectation values of observables or Born's probabilistic rule: *no reduction, no "quantum jump," no collapse of the wave packet, is needed*. This is Schrödinger's "solution" (or rather "dissolution") of the *measurement problem* of quantum mechanics.

## METAPHYSICS

Schrödinger was usually careful to separate his metaphysics from his scientific work. He held that Western science arose from the act of "objectivation"—the act of withdrawing oneself from the domain under study. By this objectivation, we push aside color, pain, esthetic judgment, and ethical values, and restrict our interest to that which is common to all: numbers and structures. But, Schrödinger argues, there is no real duality between ourselves and the objects we have thus posited. Furthermore, our personal selves are identical with the one all-comprehending universal self. Whereas science is only concerned with the relations between objectified entities, metaphysics ventures to say something about the one that comes before any objectification has taken place. This nondualist conception (which Schrödinger called the "identity theory") was overtly borrowed from the Indian *Advaita Vedānta* and was remarkably similar to Schopenhauer's earlier views. The arguments Schrödinger presents in favor of this view are as follows: (i) The truth of the "identity theory" is somehow directly experienced; (ii) The "identity theory" provides us with a coherent picture of the world as a whole, including the vexing mind-body problem; and (iii) The "identity theory" has a potentially high ethical value, because it cuts egocentrism at its root. The only point of contact between Schrödinger's metaphysics and philosophy of physics is negative. In *Mind and Matter*, Schrödinger sharply criticized Heisenberg's suggestion that quantum mechanics had weakened the Cartesian dichotomy between *res cogitans* and *res extensa*. After all, Schrödinger wrote, "Subject and object are only one. The barrier between them cannot be said to have broken down as a result of recent experience in the physical sciences, for this barrier does not exist."

**See also** Einstein, Albert; Heisenberg, Werner; Quantum Mechanics.

## Bibliography

### PRIMARY WORKS

*Science and Humanism*. Cambridge, U.K.: Cambridge University Press, 1951.

*Nature and the Greeks*. Cambridge, U.K.: Cambridge University Press, 1954.

*Science, Theory, and Man*. Mineola, NY: Dover Publications, 1957.

*My View of the World*. Cambridge, U.K.: Cambridge University Press, 1964.

*What Is Life?* and *Mind and Matter*. Cambridge, U.K.: Cambridge University Press, 1967.

*Gesammelte Abhandlungen: Verlag der österreichischen Akademie der Wissenschaften*. Vienna: Friedrich Wieweg & Sohn, 1984.

*The Interpretation of Quantum Mechanics*, edited by M. Bitbol. Woodbridge, CT: Ox Bow Press, 1995.

#### SECONDARY WORKS

Bitbol, M. *Schrödinger's Philosophy of Quantum Mechanics*. Dordrecht: Kluwer, 1996.

Moore, W. *Schrödinger: Life and Thought*. Cambridge, U.K.: Cambridge University Press, 1989.

Scott, W. T. *Erwin Schrödinger: An Introduction to his Writings*. Cambridge, MA: University of Massachusetts Press, 1967.

*Michel Bitbol (2005)*

## SCHULTZ, JULIUS

(1862–1936)

Julius Schultz, the German philosopher, dramatist, historian, and philologist, was born in Göttingen. From 1888 until 1927 he taught at a high school in Berlin. Among Schultz's numerous writings dealing with philosophy, the most important are *Die Maschinentheorie des Lebens* (1909) and *Die Philosophie am Scheidewege* (1922).

Schultz's starting point is the question How must we conceive of consciousness, on the one hand, and the object, on the other, if we wish to understand from their combined action the world of phenomena? To answer the psychological part of this question, Schultz first studied the axioms and categories of ordinary and of scientific thinking in order to see what attitude toward the phenomena is forced upon our understanding by its own innermost essence. At the same time he found a solution to the epistemological problem, namely, that if we desire not only to describe the world scientifically but also to understand it uniformly and completely, we must reduce all qualitative differences to quantitative ones. Accordingly, we must interpret the world of sense as a world of motion and explain all the happenings in the world in a mechanistic-dynamistic manner.

In epistemology, Schultz acknowledged special indebtedness to Immanuel Kant and Hans Vaihinger, whose views he interpreted and developed in a psychological fashion. His philosophy of nature is characterized

by the attempt to outline a thorough and systematic causal-mechanistic worldview. The nucleus of this view is a "machine theory of life," which Schultz developed on a broad scientific basis. The theory explains the phenomena of life with the help of the postulate of "biogenes." These are defined as unobservable molecules of submicroscopic size, which are not themselves alive but which build up the living forms. Schultz conceived of the "biogenes" in such a manner that from their joint action one can understand all the processes of life in their goal-directedness and wholeness, and thus both the forms as well as the functions of organisms. In this biomechanistic conception, organic forms are extremely complicated physicochemical systems. The goal-directed course of living processes arises out of the meaningful arrangement of these systems, and their structure and behavior are explained by strictly causal natural laws, making unnecessary the assumption of immaterial vital forces.

Schultz also contributed to the typology of philosophical thought. He sought to reduce all philosophical standpoints to two basic conceptions of the world and of life, corresponding to two different types of men. The first type pays homage to the value of conservation and prefers purposeful, useful activity; as a thinker, this practical-minded man professes an ethics of duty and believes in progress and in the efficacy of metaphysical forces. The second type prefers the value of formation and as an aesthete or theorist playfully seeks a sympathetic understanding of forms, which he desires to behold in their abundance. He professes an ethics of character, or ethics of the beauty in life, and believes in an eternal recurrence of coming into being and ceasing to be. As an advocate of determinism and causality, he envisages a mechanistic picture of the world in order to understand it in its depth. Schultz himself preferred the second standpoint, which determined his attitude in the philosophy of history. In particular, he took a pessimistic view of the future development of culture. He feared that man would become part of a machine, a socialized worker-ant—organized for common work down to the last detail, but, as in the early ages, without a history.

**See also** Consciousness; Determinism, A Historical Survey; Kant, Immanuel; Vaihinger, Hans.

#### *Bibliography*

Schultz's main philosophical works are *Psychologie der Axiome* (Göttingen, 1899); *Die Bilder von der Materie* (Göttingen, 1905); *Die drei Welten der Erkenntnistheorie* (Göttingen, 1907); *Die Maschinentheorie des Lebens* (Göttingen, 1909; 2nd ed., Leipzig: Meiner, 1929); *Die Philosophie am*

*Scheidewege* (Leipzig: Meiner, 1922); *Leib und Seele* (Berlin, 1923); *Das Ich und die Physik* (Leipzig: Meiner, 1935).

His chief historical work is *Wandlungen der Seele im Hochmittelalter*, 2 vols. (Breslau, 1936; 2nd ed., 3 vols. Breslau, 1940).

For Schultz's philosophical autobiography see *Die Philosophie der Gegenwart in Selbstdarstellungen*, edited by Raymund Schmidt (Leipzig: Meiner, 1922), Vol. III, pp. 177–198. For a complete biography and bibliography see Erwin Ditz, *Julius Schultz' Maschinentheorie des Lebens*, Vol. XIV in *Studien und Bibliographien zur Gegenwartsphilosophie* (Leipzig: Hirzel, 1935).

**Franz Austeda (1967)**

Translated by Albert E. Blumberg

## SCHULZE, GOTTLOB ERNST (1761–1833)

Gottlob Ernst Schulze, the skeptic and critic of Kantian philosophy, was born in Heldrungen, Thuringia. He was professor at Wittenberg and Helmstedt and later at Göttingen, where one of his students was Arthur Schopenhauer. His influence is due chiefly to his writings, in which he developed his critical-skeptical position. Schulze's main work, and the one that made him famous, was *Aenesidemus*. In this work, which first appeared anonymously and without the place of publication, Schulze presents objections to the Kantian critique and to K. L. Reinhold's intended vindication of the critical philosophy. Schulze's arguments against the critical philosophy led him to share David Hume's skepticism, of which he gave a concise presentation.

The *Aenesidemus* tries to show that Hume's skepticism has not been refuted by the critical philosophy. However, Schulze's position is not that of absolute skepticism: The validity of formal logic and the principles of identity and contradiction are not subject to doubt. He defined skepticism as the doctrine "that philosophy can establish neither the existence nor the non-existence of things-in-themselves and their qualities. Also the limits of our cognitive capacity cannot be fixed and ascertained on the basis of generally valid principles. ... But the reality of presentations and the certitude of mental events immediately given through consciousness no skeptic has ever doubted" (*Aenesidemus*, p. 24). On the other hand, "skepticism does not declare the metaphysical questions to be eternally unanswerable and in principle not liable to a solution" (p. 24). Through progressive development it is possible to approach a solution of the problems concern-

ing the existence or nonexistence of things-in-themselves and the limits of our cognitive capacities.

Thus the possibility of perfecting human cognition so as to attain clarity and certitude in particular metaphysical questions was not denied by Schulze. However, his objection to the critical philosophy was not limited to the question concerning the possibility of progress in metaphysics; he also attempted to show the self-contradictory nature of Immanuel Kant's critical philosophy. The critical philosophy argues that since general and necessary knowledge is possible only through synthetic a priori judgments, such judgments must represent reality. Furthermore, such judgments are possible only on the assumption of a pure capacity of understanding; hence, such a capacity must exist. In interpreting Kant, Reinhold generalized this mode of argumentation, formulating the fundamental principle that the presentation of any object implies the distinction between consciousness of the subject, of the object, and of the relation obtaining between them. From these indispensable components of the presentation Reinhold concluded the reality of corresponding objects.

However, from the fact that presentations always contain the notions of subject, of object, and of their relation to each other it is illegitimate, according to Schulze, to conclude the objective reality of corresponding objects. The transition from thought to being is grounded in ontological thinking, which Kant himself showed to be defective in his criticism of the classic proofs for the existence of God and of dogmatic metaphysics. Since one cannot argue from the conditions of thought to the reality of objects, the problem of philosophy is, according to the critical philosophy, to search for the competence and the legitimacy of our thought to determine objects of reality. The task of the Kantian critique is to show the objective validity of our judgments. However, the indispensable conditions of thought constitute subjective necessity, from which objective validity cannot be derived.

Furthermore, "it is presupposed that each part of human cognition must be grounded in reality as its cause. Without such an assumption the doctrine of the *Critique* concerning the origin of the necessary judgments has no meaning whatsoever" (*ibid.*, pp. 137f.). The conclusion from the necessary judgments in our consciousness as to the reality of objects is based on the principle of causality. Existing objects constitute the causes of our cognition. The category of causality is thus employed with reference to noumena. Also, in the conception of sensibility as a faculty of receptivity, the existence of things-in-them-



selves that have the capacity to affect our sensibility is presupposed. Here again the concept of causality is applied to noumena, while, according to the critical philosophy, causality as a category of understanding is confined to the realm of phenomena. Reinhold's doctrine that things-in-themselves, although not cognizable, are nonetheless thinkable, is untenable. Since the things-in-themselves are thought to be the cause of cognition, they are cognized as having the capacity to affect the knowing and thinking capacity. The thing-in-itself must be cognizable, or it cannot be considered as a cause of cognition.

Likewise, the concept of causality cannot be employed for proving the reality of the subject as a thing-in-itself. Schulze understood the Kantian solution of the question "How are synthetic a priori propositions possible?" as consisting in the derivation of these propositions from the subject as their cause: "The *Critique* derives the necessary synthetic propositions from the subjective mind (*Gemüth*) and its a priori determined cognitive processes ... by the application of the principle of causality, which does not harmonize with its own principles delimiting the area of application of the categories" (ibid., pp. 153f.). Moreover, the conclusion from the propositions to the reality of a capacity in the mind does not explain the process of cognition. Nothing is gained by proposing that the perception of the material given is due to a receptive capacity, for a problem is not explained by reducing it to something unknown. The problem of cognition of experience is not solved by a reduction of cognition to a receptive capacity that is no less problematical.

Schulze considered the a priori concepts as existing in time "prior" to the cognition of objects. This account of the a priori concepts as innate ideas and as inherent qualities of the subjective mind is a misunderstanding of the Kantian position that has been common to numerous interpreters of Kant until the present. Schulze thus failed to understand the essence of the critical philosophy, which does not aim at deriving the synthetic propositions from the subject as a thing-in-itself. Kant was not concerned with the psychological process of cognition but with objective cognition, as manifested in the scientific process. The problem is how synthetic a priori propositions are possible in mathematics and science, and not how the human mind as a subject conceives such propositions. The objectivity of the judgments is vouchsafed by the scientific laws determining objects that arise through these laws. This is implied in the Kantian principle of the "possibility of experience." Scientific experience is possible only through synthetic propositions. Since without

synthetic propositions there would be no scientific experience at all, their legitimacy is vouchsafed by the function they fulfill for experience. Furthermore, Schulze held the difference between synthetic and analytic propositions was not an objective distinction, and, psychologically considered, it depends on subjective circumstances whether a proposition is synthetic or analytic for a particular individual at a certain moment.

Schulze's criticism of Kant implied the notion of the subject and predicate of the proposition as individually given and fixed entities, so that the synthetic proposition connects elements that can be thought of in themselves. Hence, the concepts of the subject and the predicate must be thought of as separately given, and the question is how their connection can be of a necessary nature. Schulze did not realize that for Kant the concept of the subject arises by its determination through the synthetic proposition. In the proposition "*S is P*," *S* is an unknown before its determination through *P*. The investigation of *S* is a "doubt-inquiry process" (John Dewey's expression); *S* acquires determination only through the predicate. Schulze's criticism is thus predicated upon an understanding of the critical philosophy as subjective idealism with the notion of a priori concepts as innate ideas, which leads to dogmatic assumptions concerning the application of the concept of causality to things-in-themselves. The a priori concepts in the critical philosophy are not to be understood as constituent features of the subjective human mind but as creative functions of thought in the process of ordering experience.

Schulze was also critical of Kant's conception of moral theology. He raised objections to the Kantian doctrine of the postulates (God, freedom, immortality) as formulated in the *Kritik der praktischen Vernunft*. From the sense of the moral command in us, the categorical imperative, there can be no conclusion as to the reality of a most perfect being. As ideas of reason, God, freedom, and immortality are endless tasks for human activity, but by the conception of these ideas as postulates their real existence as objects is posited. "The Kantian moral theology postulates *more* than what practical reason demands for the satisfaction of its requirements" (ibid., pp. 440ff.). In his criticism of the postulates Schulze has partly anticipated the neo-Kantianism of the Marburg school. Hermann Cohen, for example, although motivated by different considerations, has pointed out that the regulative ideas of reason do not require the support of the doctrine of postulates.

Schulze's contribution to the development of Kantian idealism consists in his exposing the contradictions

and inconsistencies involved in both dogmatic-realistic and subjective-idealistic interpretations of the critical philosophy, but his attempt at a vindication of Hume's skepticism proved ineffective for further development of Kantian idealism. Philosophical thought took the course not back to Hume but to a more consistent critical idealism eliminating the concept of a thing-in-itself (as in Salomon Maimon) and to speculative idealism as it developed in the post-Kantian metaphysical systems. However, by his valuable criticism of the doctrine of the faculties of the soul Schulze anticipated Johann Friedrich Herbart and influenced Friedrich Eduard Beneke (1798–1854).

According to Schulze, a phenomenon of the life of the soul is not explained by attributing it to a "faculty." Such an attribution does not explain, but merely gives another name to the same thing. The task of psychology as a science is, rather, a detailed description of actual mental occurrences and their systematic classification. By such a method, general concepts of psychological phenomena can be attained; but they should not be attributed to "faculties" of the soul, which is a metaphysical concept.

**See also** A Priori and A Posteriori; Beneke, Friedrich Eduard; Cohen, Hermann; Dewey, John; Herbart, Johann Friedrich; Hume, David; Innate Ideas; Kant, Immanuel; Maimon, Salomon; Neo-Kantianism; Reinhold, Karl Leonhard; Skepticism.

### Bibliography

#### WORKS BY SCHULZE

*Grundriss der philosophischen Wissenschaften*. 2 vols. Wittenberg, 1788–1790.

*Aenesidemus, oder über die Fundamente der von dem Herrn Professor Reinhold in Jena gelieferten Elementarphilosophie. Nebst einer Vertheidigung des Skepticismus gegen die Anmassungen der Vernunftkritik*. 1792. Reedited by the Kantgesellschaft, in the series of rare philosophical works edited by Arthur Liebert. Berlin: Reuther and Reichard, 1911.

*Kritik der theoretischen Philosophie*. 2 vols. Hamburg, 1801.  
*Enzyklopädie der philosophischen Wissenschaften*. Göttingen, 1814.

*Über die menschliche Erkenntnis*. Göttingen, 1832.

#### WORKS ON SCHULZE

Cassirer, Ernst. *Das Erkenntnisproblem in der Philosophie und Wissenschaft der neueren Zeit*. Vol. III. Berlin: Cassirer, 1920.

DiGiovanni, George. "The First Twenty Years of Critique: The Spinoza Connection" In *The Cambridge Companion to Kant*, edited by Paul Guyer. New York: Cambridge University Press, 1992.

Erdmann, Johann Edward. *Versuch einer wissenschaftlichen Darstellung der Geschichte der neueren Philosophie*. Vol. III. Riga and Dorpat, 1848–1853.

Fischer, Kuno. *Geschichte der neueren Philosophie*. 3rd ed., vol. VI. Heidelberg, 1900.

Kroner, Richard. *Von Kant bis Hegel*. 2 vols. Tübingen: Mohr, 1921–1924.

Rosen, Michael. "From Kant to Fichte: A Reply to Franks" In *Transcendental Arguments: Problems and Prospects*, edited by Robert Stern. Oxford: Clarendon Oxford Press: 1999.

Wieggershausen, Heinrich. *Aenesidem-Schulze, der Gegner Kants und seine Bedeutung im Neukantianismus*. Berlin, 1910. This is *Kantstudien*, Supp. No. 17.

**Samuel Atlas (1967)**

*Bibliography updated by Tamra Frei (2005)*

## SCHUPPE, ERNST JULIUS WILHELM

(1836–1913)

Ernst Julius Wilhelm Schuppe, the German philosopher, was born in Brieg, Silesia. He studied at the universities of Breslau, Bonn, and Berlin, and he took his doctorate at Berlin in 1860. He taught at grammar schools in Silesia and then held a chair of philosophy at the University of Greifswald from 1873 to 1910.

### EPISTEMOLOGY

In his main work, *Erkenntnistheoretische Logik* (Bonn, 1878), largely anticipated by his earlier book *Das menschliche Denken* (Berlin, 1870) and summarized in his later *Grundriss der Erkenntnistheorie und Logik* (Berlin, 1894; 2nd ed., Berlin, 1910), Schuppe was concerned with the epistemological bases of knowledge generally and of logic in particular. Schuppe held that a theory of knowledge should avoid hypotheses such as the transcendent reality postulated by realists and metaphysicians, but that it should equally avoid one-sided objective or subjective foundations of knowledge, whether materialist, positivist, or idealist.

In keeping with these requirements, Schuppe developed the notion of conscious immanence (*Immanenz, Bewusstsein, Ich*) in which subject and object form a unity. This immanence of consciousness, or ego, is a fact (*Tatsache*) that is given with certitude and can therefore serve as a starting point for epistemology. Only abstractly can the ego be divided into subject and object; concretely it is a correlation of the two. This is not to say that the object is a psychic entity, but merely that there is no being not related to a subject. To ignore the correlation would

be to incur a contradiction because a supposed unthought entity is nevertheless implied in the thought of the epistemologist.

To account for the distinction and division of consciousness and content (*Inhalt*), and of contents among themselves (subjective elements such as acts are distinct from objects of acts, however much both may have to be considered contents for an abstract subject), Schuppe presented a theory of “common” content: Objective content is a given that can be shared by several, whereas subjective content (sensation, for example) is unique and private. The need for this division led to a theory of consciousness in general (*Bewusstsein überhaupt*) as distinguished from the consciousness of a concrete individual subject. Individuality is based on content not shared by others. Other minds, which are presupposed by the notion of consciousness in general, are known, Schuppe claimed, by inference mediated by one’s own body; but he also asserted that they can be regarded as immediately perceived. Schuppe denied the claim that other minds are immanent contents of one’s mind—like any other object—as being tantamount to solipsism. Schuppe drew upon the ontic fact of a plurality of minds as a basis for consciousness in general.

Schuppe held that thought is also a “component” of the content of consciousness, along with the sense component; it “accedes” to perceptual data. Accordingly, objects of cognition can be considered as constituted by an interaction of an original given of sense, by itself an abstraction, with performances of thought (*Denkarbeit*). In fact Schuppe came to regard thought as the central function of consciousness: To think is to appropriate content, to receive an impression in its positive determinacy, to fixate it as identical. This primary performance of appropriation is thought-in-general, which is prior to judgment. Schuppe argued that at this stage there is only one datum to be appropriated but that for judgment two contents, subject and predicate, are required. (Here Schuppe was influenced by a grammatical notion of judgment.) Continuing to develop his notion of content, Schuppe introduced an analysis of content in which thought stands for the identification of two contents (an instance of the principle of identity, with the principles of contradiction and limitation as corollaries) and, somewhat surprisingly, for the establishment of causal connection between them. Identity and causality are the categories that constitute objective content. (Here Schuppe was guided by a metagrammatical or transcendental notion of judgment, interpreting the category as the predicate of the unified contents.)

## ONTOLOGY

With this basis of transcendental thought, Schuppe’s “epistemological” logic was not so much concerned with the “forms” of formal logic as with the establishment of a priori truths about the object of knowledge. Thus the logic constitutes a theory of objects, an ontology. Schuppe analyzed the given into its elements (temporal and spatial determinateness, sense impression) and conceptual moments (genera and species), and distinguished several kinds of union (*Zusammengehörigkeit*) among them. In a transcendental progression Schuppe established number, space region, thing, organism, and artifact; and genera, species, and matter. He avoided any reference to a transcendent cause. Understandably, he presented a coherence theory of truth.

## LOGIC

Schuppe sought a transcendental genealogy as a basis for logic. This project involved a certain deviation from the traditional understanding of formal logic. He rejected the isolation in logic of a purely formal realm, denying in fact that purely formal theorems are significant. He regarded propositions as assertions of categorial unification. Logic must be concerned with the realm of material content in which unity is asserted and must examine the various types of union of content, that is, the “real” genera of content, which, in the case of objects of appearance, are grounded in the causal context. This doctrine has ramifications in many areas of logic, for example, in the theory of definition.

Schuppe’s theoretical philosophy can be regarded as a doctrine of the constitution of knowledge and its objects by transcendental synthesis. In view of its intuitive starting point and its analysis of given content, however, it seems to be a compromise between a logico-transcendental theory and a theory of reflective intuition. The agency responsible for the grounding of objectively constituted content is both a transcendental principle and an existent consciousness. The normative element of a transcendental theory is merged with the factual basis of a subjective ontology. Schuppe’s philosophy thus stands between transcendental critique and ontological philosophy of immanence. Although it leans heavily on Immanuel Kant, it anticipates much of Edmund Husserl’s phenomenology and constitutes an example for a theoretical understanding of the interplay of factuality and logico-transcendental thought.

## PRACTICAL PHILOSOPHY

Schuppe's *Grundzüge der Ethik und Rechtsphilosophie* (1881; reprinted 1963) offers an independent compromise between a normative position, based on the will as a form of consciousness in general, and a eudaemonistic one, based on pleasure. He also wrote several studies in the philosophy of law, such as *Der Begriff des subjektiven Rechts* (Breslau, 1887), and joined the philosophical discussion concerning the new German civil code (*Das Gewohnheitsrecht*, Breslau, 1890; *Das Recht des Besitzes*, Breslau, 1891).

**See also** Coherence Theory of Truth; Epistemology; Epistemology, History of; Husserl, Edmund; Kant, Immanuel.

### Bibliography

In addition to the works by Schuppe already cited, see *Allgemeine Rechtslehre mit Einschluss der allgemeinen Lehren vom Sein und vom Wissen*, edited and with an introduction by Wilhelm Fuchs (Berlin: Verlag für staatswissenschaften und geschichte, 1936).

For works on Schuppe, see R. Hermann, *Schuppes Lehre vom Denken* (Greifswald, 1894); Gunther Jacoby, *Wilhelm Schuppe*, no. 45, *Greifswalder Universitätsreden* (Greifswald, 1936); L. Kljubowski, *Das Bewusstsein und das Sein bei Wilhelm Schuppe* (Heidelberg, 1912); Rudolf Zocher, *Husserls Phänomenologie und Schuppes Logik* (Munich, 1932).

*Klaus Hartmann (1967)*

## SCHUTZ, ALFRED

(1899–1959)

Alfred Schutz was born in Vienna on April 13, 1899. He studied law and social sciences at the University of Vienna from 1918 until 1921, where he completed a doctorate in law and then continued his studies in the social sciences until 1923. Equally as important for his intellectual development as his studies at the university was his participation in the informal academic life of Vienna, in which he also cultivated his philosophical interests. After completing his academic studies and in addition to his ongoing scholarly activities, Schutz held the full-time job of a bank lawyer—a dual life that lasted until he retired from the bank in 1956. Following the annexation of Austria to the Third Reich, Schutz and his family escaped via Paris to New York. There he became affiliated with the graduate faculty at the New School for Social Research in New York, where he taught from 1943 until 1952 as a lecturer, then afterwards as a full professor of sociology and from

1956 as professor of both sociology and philosophy. He died on May 20, 1959, in New York.

Schutz is regarded as the founder of the phenomenological approach in sociology. Influenced by Max Weber, Henri Bergson, Edmund Husserl, the Austrian School of Economics, and pragmatism, he sought to give a philosophical foundation to interpretative social sciences. As a critical follower of Husserl, he developed his own *mundane phenomenology* of the life-world and its structures, showing how actors produce and understand social reality in everyday interactions and communication.

Schutz begins with Max Weber's view of social reality as a meaningful sociocultural world and shares his concept of meaning-oriented social action, but he criticizes Weber for neglecting to inquire into the constitution of meaning in general. In order to analyze how the meaning attached to action is revealed, Schutz refers to the philosophical concepts developed by Henri Bergson and Edmund Husserl. He adopts the Bergsonian idea of the stream of consciousness, but he later comes to recognize difficulties in Bergson's intuitivism and turns to Husserl's phenomenology.

In 1932 Schutz writes his masterpiece “Der sinnhafte Aufbau der sozialen Welt” where he develops his basic concept of the constitution of the social world and formulated his own phenomenological position. Influenced by pragmatism—which was mediated to him by Henri Bergson, Max Scheler, and William James—in this work he proceeds beyond the realm of consciousness and perception as analyzed by Husserl and considers both human action and interactions as well as acts of consciousness as factors in the constitution of meaning. Leaving behind the transcendental philosophical approach, he develops his own “mundane” phenomenology that analyzes the constitution of a meaningful world within “mundane” social relationships in the everyday world. He adopts the results of the Husserlian analyses of the temporality of consciousness, of the intentional structure of lived experience, as well as of the meaning constitution based on embodiment as preconditions on which the social shaping of experience patterns is based, but he rejects Husserl's assumption that intersubjectivity in the sense of understanding of the others could be a result of the acts of consciousness alone (Schutz 1966). Rather, he shows how the schemes of experience are shaped by influencing (*Wirken*) and by relationships of influence (*Wirken-beziehung*) which consist of interaction and communication.

Schutz understands communication as a process where two subjective streams of consciousness are coordinated within a relationship of mutual influence and where the meaning of one ego's action consists in the intention to evoke a reaction on the part of the other. Actions have here the function of signs that are mutually indicated and interpreted. Because the final meaning of one person's action is revealed in the reaction of the other and vice versa, communication provides a common stock of shared patterns of interpretation that allows for mutual understanding, even if each of the agents always refers to his or her own schemes of experience. In this concept of understanding based on interaction, Schutz offers his own solution to the problem of intersubjectivity posed by Husserl.

In his later work Schutz (1962, 1964, 1966) determines this communicatively created social reality as the world of everyday life whose typical patterns are taken for granted and represent the intersubjective common core of the reality in which people live. He also discloses further structural characteristics of this everyday core of the life-world: Its typical structure depends considerably on the pragmatic orientation of action selecting the areas where typification processes take place (1962, 1966, 1970). Both typicality and this selection based on *systems of relevance* represent two generative principles of order in the everyday world. This everyday reality is nevertheless not identical to the life-world as a whole. By suspending his pragmatic interest, the agent is able to modify his or her everyday experiences and perceive them as objects of a game, fantasy, art, science, or as a dream. All those modifications represent different provinces of meaning that transcend the everyday world and constitute the *multiple realities* (Schutz 1962) of which the life-world is composed. The different strata of meaning in the life-world are integrated by semiotic systems whose structure allows the contents of one province of meaning to be symbolized by another through appresentation.

By considering communication as a substantial constitutive mechanism of social reality, Schutz (1962) stresses the role of language in this process. On his view, language maintains relevances and typifications unique to specific cultures and to social groups and is thus crucial for the constitution of the life-world as a cultural one (Schutz and Luckmann 1989).

The methodological rule that Schutz derives from his approach is expressed in his *postulate of adequacy* (Schutz 1962, 1964) between everyday and scientific typifications. This postulate holds that higher-order interpretative types employed by social sciences have to be constructed

in correspondence to the structure of the everyday typifications (first-order types). Thus the structure of the life-world that guides everyday actions also represents the methodological framework within which the social and cultural sciences have to proceed.

The Schutzian phenomenological approach represents one of the main paradigms in the area of interpretative social and cultural sciences. In philosophy his theory led to a critical assessment of the Husserlian view of intersubjectivity, to conceptions of a worldly phenomenology and theory of the cultural sciences (Embree 1988), and to a philosophy of modern anonymity (Natanson [1962–1995] 1986), as well as to new insights into intercultural hermeneutics (B. Waldenfels 1997, 1998, 1999). It also influenced the philosophy of gender (E. List 1993). In a modified form, Schutz's concept of the life-world was also integrated into the social philosophy of Jürgen Habermas.

*See also* Phenomenology.

## Bibliography

### WORKS BY SCHUTZ

- Der sinnhafte Aufbau der sozialen Welt*. Wien, Austria: Springer, 1932. Translated by George Walsh and Frederick Lehnert as *The Phenomenology of the Social World* (Evanston, IL: Northwestern University Press, 1967).
- Collected Papers*. 4 vols., edited by Maurice Natanson. The Hague: M. Nijhoff, 1962–1995. Vol. 1, *The Problem of Social Reality* (1962); vol. 2, *Studies in Social Theory*, edited by A. Broderson (1964); vol. 3, *Studies in Phenomenological Philosophy*, edited by I. Schutz (1966); vol. 4, [without special title], edited by Helmut Wagner and George Psathas with Fred Kersten (1995).
- Reflections on the Problem of Relevance*, edited by Richard M. Zaner. New Haven, CT: Yale University Press, 1970.
- The Structures of the Life-World*. With Thomas Luckmann. Evanston, IL: Northwestern University Press, 1973. Revised edition, 1989.
- Life Forms and Meaning Structure*. Translated by Helmut Wagner. London: Routledge, 1982.
- Werkausgabe*. 12 vols. Konstanz, Germany: Un[iv]ersitätsverlag Konstanz, 2003–.

### WORKS ON SCHUTZ

- Barber, Michael D. *The Participating Citizen: A Biography of Alfred Schutz*. Albany: State University of New York Press, 2004.
- Embree, Lester, ed. *Worldly Phenomenology: The Continuing Influence of Alfred Schutz on North American Human Science*. Washington, DC: University Press of America, 1988.
- List, Elisabeth. *Die Präsenz des Anderen: Theorie und Geschlechterpolitik*. Frankfurt am Main, Germany: Suhrkamp, 1993.
- Natanson, Maurice. *Anonymity: A Study in Philosophy of Alfred Schutz*. Bloomington: Indiana University Press, 1986.

Srubar, Ilija. *Kosmion: Die Genese der pragmatischen Lebensweltheorie von Alfred Schütz und ihr anthropologischer Hintergrund*. Frankfurt am Main, Germany: Suhrkamp, 1988.

Wagner, Helmut R. *Alfred Schütz: An Intellectual Biography*. Chicago: University of Chicago Press, 1983.

Waldenfels, Bernhard. *Studien zur Phänomenologie des Fremden*. 4 vols. Frankfurt am Main, Germany: Suhrkamp, 1997, 1998, 1999.

*Ilija Srubar (2005)*

## SCIACCA, MICHELE FEDERICO (1908–1975)

Michele Federico Sciacca was a founder of the Gallarate movement, professor of theoretical philosophy at the University of Genoa, and the founder and editor of the journals *Giornale di metafisica* and *Humanitas*. He started as a historian of ideas, writing important works on Reid (1935), Plato (1939), and St. Augustine (1939); a massive review of Italian thought, *Il XX secolo* (2 vols., Milan, 1941); and a review of contemporary European thought, *La filosofia oggi* (Milan, 1945).

Although Sciacca studied under Antonio Aliotta, his major stimulus came from Giovanni Gentile, from whom Sciacca derived his basic axiom that concrete being must be act, never fact. Sciacca developed this principle in his own fashion under the influence of Plato, St. Augustine, Antonio Rosmini-Serbati, and Maurice Blondel.

Sciacca's position was one of "integralism." The central notion of integralism is interiority, according to which the ground of all forms of being and existence lies in the activity of the subject. Sciacca asserts that the existent, or act, cannot be a fact among facts; its existence resides wholly in its own self-generative actuality. Against existentialism he asserts that the being of the existent cannot be pure possibility or nothingness; it must be being. The whole concern of integralism is to establish the character of the being that the existent is. Sciacca holds this being to be objective interiority, which he delineates in his most original speculative work, *Interiorità oggettiva* (Milan, 1951). Interiority is the positing by the existent of itself as act. So defined, it cannot be conceived as purely immanent, in the manner of Gentile. It must posit itself with reference to a transcendent and objective reality and define itself within this horizon. The basic structural principle of interiority is truth, or the subject's affirmation of the ground of its existence in the very act of existing. The immanent ground of the subject and of all

existence is a transcendent being, not abstract but more concrete and existentially real than the subject—God. In affirming the existence of God, the subject also affirms its own being, the innermost character of its own act of existing.

Sciacca's basic insight is thus that the being of the subject cannot be mere possibility, nothingness, or facticity but must be act; that this act is the affirmation of its own actuality through the affirmation of its transcendent ground; and that the absolute existent is present in concrete human existence. It is this presence of the Absolute that establishes the human existent as a person. In *Morte ed immortalità* (Brescia, 1954), Sciacca holds that the affirmation of God within human existence that constitutes the human subject cannot be a merely transitory relationship and that immortality is therefore the logical extension of interior objectivity.

*See also* Gentile, Giovanni.

### Bibliography

Works on Sciacca include the following: Sante Alberghi, "Sul pensiero di M. F. Sciacca," in *Giornale critico della filosofia italiana*, Series III, 4 (1950): 841–893; A. Robert Caponigri, "The Existence of God: Proof from Truth," in *Saggi in onore di M. F. Sciacca* (Milan, 1960), pp. 53–78; Cornelio Fabro, "La verità integrale dell'uomo integrale," in *Divus Thomas*, Series III, 27 (1950): 511–519; Augusto Guzzo, Romeo Crippa, and A. Arata, eds., *Michele Federico Sciacca* (Turin: Edizioni di "Filosofia," 1951), which contains a complete bibliography of Sciacca's writings and of writings about him; and "Note critiche sul neo-spiritualismo," in *Giornale di metafisica* 9 (2) (1956), which contains contributions by various authors.

### WORKS BY SCIACCA

*L'interiorità oggettiva*. 4th ed. Milano: Marzorati, 1965.  
*La libertà e il tempo*. Milano: Marzorati-editore, 1965.  
*La filosofia e la scienza nel loro sviluppo storico. Per i licei scientifici*. Roma, Cremonese, 1965–1967.  
*Dallo spiritualismo critico allo spiritualismo cristiano (1939–1951)*. Milano: Marzorati, 1965.  
*Lettere dalla campagna e scritti affini*. 2nd ed. Milano: Mazarati, 1966.  
*Opere complete. Saggi, scritti inediti, opere minori, recensioni*. Milano: Marzorati 1966.  
*Lettere dalla compagna e scritti affini*. 2nd ed. Milano: Marzorati, 1966.  
*Platone*. 2nd ed. Milano: Marzorati, 1967.  
*Il Pensiero moderno (secoli XVII–XVIII)*. Milano: Marzorati, 1968.  
*Morte e immortalità*. 3rd ed. Milano: Marzorati, 1968.  
*Studi sulla filosofia moderna*. 4th ed. Milano: Marzorati, 1968.  
*Filosofia e antifilosofia*. Milano: Marzorati, 1968.  
*La filosofia morale di Antonio Rosmini*. 5th ed. Milano: C. Marzorati, 1968.

*Il problema di Dio e della religione nella filosofia attuale*. 5th ed. Milano: Marzorati, 1968 .

*In spirito e verità: pensieri e meditazioni*. 7th ed. Milano: Marzorati, 1969.

*Pagine di critica letteraria*. (1931–1935). Milano: Marzorati, 1969 .

*Autobiografia e risposte al Giornale de' letterati d'Italia*. Milano: Marzorati, 1969.

*Gli arieti contro la verticale*. Milano: Marzorati, 1969.

*La Chie sa e la civiltà moderna*. 2nd ed. Milano: Marzorati, 1969.

*Cultura e anticultura*. Torino: Borla, 1969 .

*La filosofia, oggi*. Milano: Marzorati, 1970.

*L'oscuramento dell'intelligenza*. Milano: Marzorati, 1970.

*Interpretazioni Rosminiane*. 3rd ed. Milano: Marzorati, 1971.

*Dialogo con Maurizio Blondel*. 2nd ed. Milano: Marzorati, 1971.

*Studi sulla filosofia antica. Con un'appendice sulla filosofia medioevale*. 2nd ed. Milano: Marzorati, 1971.

*Il chisciotismo tragico di Unamuno e altre pagine spagnole*. Milano: Marzorati, 1971.

*Ontologia triadica e trinitaria*. Stresa: Centro internazionale di studi rosminiani, 1972.

*Pagine di pedagogia e di didattica*. Milano: Marzorati, 1972.

*Il pensiero italiano nell'età del risorgimento*. 3rd ed. Milano: Marzorati, 173.

*Pascal*. 8th ed. Milano: Marzorati, 1973.

*L'ora di Cristo*. 2nd ed. Milano: Marzorati, 1973.

*L'estetismo, Kierkegaard, Pirandello*. Milano: Marzorati, 1974.

*Pirandello, I giganti della montagna*. Stresa: Centro internazionale di studi rosminiani, 1974.

*Figure e problemi del pensiero contemporaneo*. Milano: Marzorati, 1974

*Prospettiva sulla metafisica di San Tommaso*. Milano: Città Nuova editrice, 1975.

*Il magnifico oggi*. Roma: Città nuova, 1976.

*La filosofia di M.F. Sciacca: saggi*. Bologna: Tip. editrice compositori, 1976.

*Il pensiero della rinascenza e della riforma: protestantesimo e riforma cattolica*. Milano: Marzorati, 1977.

*L'interiorità oggettiva*. Palermo: L'epos, 1989.

*Il chisciotismo tragico di Unamuno*. Palermo: L'epos, 1989.

*Pascal*. Palermo: L'epos, 1989.

*Dialogo con Maurizio Blondel*. Palermo: L'epos, 1989.

*La filosofia morale di Antonio Rosmini*. Palermo: L'epos, 1990.

*Platone*. Palermo: L'epos, 1990.

*L'estetismo, Kierkegaard, Pirandello*. Palermo: L'epos, 1990.

*Atto ed essere*. Palermo: L'epos, 1991.

*Sant'Agostino*. Palazzo Speciale: Epos, 1991.

*Prospettiva sulla metafisica di S. Tommaso*. Palermo: L'epos, 1991.

*La filosofia italiana nel secolo XX*. Palermo: L'epos, 1997–.

*Filosofia e antifilosofia*. Palermo: L'epos, 1998.

*Come si vince a Waterloo*. Palermo: L'epos, 1999

*La casa del pane*. Palermo: L'epos, 1999.

*La libertà e il tempo*. Palermo: L'epos, 1999.

*Gli arieti contro la verticale*. Palermo: L'epos, 1999.

*L'uomo questo squilibrato*. Palermo: L'epos, 2000.

*L'oscuramento dell'intelligenza*. Palermo: L'epos, 2000.

*Filosofia e metafisica*. Palermo: L'epos, 2002.

*I principi della metafisica rosminiana*. Palermo: L'epos, 2003.

**A. Robert Caponigri (1967)**

*Bibliography updated by Michael J. Farmer (2005)*

## SCIENCE, RESEARCH ETHICS OF

The idea that ethics is important in scientific research is not new. In 1830 Charles Babbage (1791–1871) admonished British scientists for engaging in dishonest research. In 1912 researchers discovered the fossil skull of a missing link between humans and apes at the Piltdown quarry in Sussex, England. After four decades of controversy, several scientists proved that the skull was a hoax.

At the beginning of World War II, prominent physicists believed that it was their moral obligation to help defeat Nazi Germany. Albert Einstein wrote a letter to President Franklin D. Roosevelt (1882–1945) urging the United States to develop the atomic bomb. J. Robert Oppenheimer (1904–1967) directed the Manhattan Project, a \$1 billion effort to build the first nuclear weapons. After the United States dropped two bombs on Japan in the summer of 1945, many scientists who worked on the bomb also led the Atoms for Peace movement, which helped to establish the International Atomic Energy Commission. During the Nuremberg Trials (1949–1949), the international community adopted a code of conduct for human experimentation, the Nuremberg Code (1947), in response to the horrific experiments on human subjects conducted by Nazi researchers at Nuremberg.

In 1961 Rachel Louise Carson (1907–1964) alerted the public to the toxic effects of the pesticide DDT on animal species and helped to launch the environmentalist movement. In 1966, Henry Knowles Beecher (1904–1976) published an article describing twenty-two ethically problematic medical experiments, including the Tuskegee Syphilis Study. In this experiment, which took place from 1932 to 1972 in Tuskegee, Alabama, researchers withheld medical treatment from African American subjects with advanced syphilis, even after an effective treatment, penicillin, became available in the 1940s. The study continued until the media brought it to the public's attention in 1972, prompting Congress to hold hearings on biomedical research and adopt new laws pertaining to research on human subjects. In 1975, philosopher Peter Singer published a book that challenged the moral legitimacy of most experiments on animals and helped to energize the growing animal rights

movement. In that same year scientists held a conference at Asilomar, California, on the risks of genetically engineered microorganisms.

Interest in the ethics of research increased dramatically in the mid-1980s due to at least two factors. First, there were many highly publicized allegations of data fabrication (making up data), falsification (changing data), and other unethical activities in federally funded research. Second, the academic research enterprise became much more commercialized due to changes in intellectual property laws and the expansion of the pharmaceutical and biotechnology industries. In the 1980s, patent offices began awarding patents on many different biological products and processes, such as DNA, cell lines, and genetically modified organisms. The U.S. Congress also passed several laws encouraging the transfer of technology from the public to the private sector. In response to these changes in the law, universities began aggressively pursuing and protecting intellectual property. Academic researchers also took a greater interest in intellectual property and in forming start-up companies to commercialize new inventions and discoveries. The pharmaceutical industry increased its spending on research and development, and the biotechnology industry, which emerged in the late 1970s following the development of gene sequencing, splicing, and copying techniques, did the same. By the beginning of the twenty-first century, private industry accounted for more than sixty percent of all research and development expenditures in the United States.

As research became more commercialized, financial ties between academic and government scientists and private companies, for example, ownership of stock or patents and gifts or consulting arrangements, increased. These financial interests created a conflict of loyalties for scientists and universities and threatened the objectivity and trustworthiness of research. Scientists, ethicists, and journalists presented evidence that some researchers and private companies were biasing data analysis and interpretation, research design, and publication practices to produce results favorable to those companies. Financial interests (and pressures) in research also were linked to fabrication, falsification, and other ethical problems.

From the late-1980s to the early twenty-first century, many different organizations took steps to promote ethics in research. The National Institutes of Health (NIH) mandated that all students on Public Health Service training grants and all intramural researchers receive instruction in responsible conduct of research. Universities incorporated ethics education into the graduate cur-

riculum to meet NIH requirements and to minimize the risk of the legal liability and public embarrassment from ethical misconduct in research. The NIH and the National Science Foundation adopted a common definition of *research misconduct* as well as policies and procedures for investigating and adjudicating misconduct allegations. The National Academy of Sciences (NAS) published several reports concerning ethics in research. Many different professional organizations and scientific journals adopted or revised codes of conduct in research.

Research ethics has become multidisciplinary field of scholarship, education, and policy, encompassing the humanities, the social sciences, and the natural sciences. Some of the key topics in the field include: the foundations of research ethics; ethical decision-making in research; recording, storing, and sharing data; honesty and objectivity in research; scientific misconduct; authorship and publication; collaboration; mentoring; intellectual property; ownership of research materials; conflicts of interest; diversity in science; research on human and animal subjects; research in genetics and biotechnology; scientific freedom; social responsibility in research; research funding; and legal and regulatory aspects of research.

**See also** Einstein, Albert; Philosophy of Social Sciences; Singer, Peter.

### **Bibliography**

- "About ORI." Historical Background. Office of Research Integrity. Available from [www.ori.dhhs.gov/](http://www.ori.dhhs.gov/).
- Angell, Marcia. *The Truth about the Drug Companies: How They Deceive Us and What to Do about It*. New York: Random House, 2004.
- Babbage, Charles. *Reflections on the Decline of Science in England, and Some of Its Causes* (1830). New York: Augustus Kelly, 1970.
- Beecher, Henry. "Ethics and Clinical Research." *New England Journal of Medicine* 274 (1966): 1354–60.
- Broad, William, and Nicholas Wade. *Betrayers of the Truth: Fraud and Deceit in the Halls of Science*. New York, Simon and Schuster, 1983.
- Brody, Baruch. *The Ethics of Biomedical Research: An International Perspective*. New York: Oxford University Press, 1998.
- Bulger, Ruth Ellen, Elizabeth Heitman, and Stanley Joel Reiser, eds. *The Ethical Dimensions of the Biological Sciences*. New York: Cambridge University Press, 2002.
- Carson, Rachel. *Silent Spring*. Boston: Houghton Mifflin, 1962.
- Dresser, Rebecca. *When Science Offers Salvation: Patient Advocacy and Research Ethics*. New York: Oxford University Press, 2001.
- Emanuel, Ezekial, Robert Crouch, John Arras, et al. *Ethical and Regulatory Aspects of Clinical Research: Readings and*



- Commentary. Baltimore, MD: Johns Hopkins University Press, 2003.
- “Guidelines for the Conduct of Research in the Intramural Research Programs at NIH.” Available from [www.nih.org](http://www.nih.org).
- Hart, Lynette A., ed. *Responsible Conduct with Animals in Research*. New York: Oxford University Press, 1998.
- Institute of Medicine. *Integrity in Scientific Research: Creating an Environment that Promotes Responsible Conduct*. Washington, DC: National Academy Press, 2002.
- Jones, Anne, and Faith McLellan, eds. *Ethical Issues in Biomedical Publication*. Baltimore, MD: Johns Hopkins University Press, 2000.
- Jones, James. *Bad Blood: The Tuskegee Syphilis Experiment*. New York: Free Press, 1981.
- Kevles, Daniel. *The Baltimore Case: A Trial of Politics, Science, and Character*. New York: Norton, 1998.
- Krimsky, Sheldon. *Science in the Private Interest: Has the Lure of Profits Corrupted Biomedical Research?* Lanham, MD: Rowman & Littlefield, 2003.
- LaFollette, Hugh, and Niall Shanks. *Brute Science: Dilemmas of Animal Experimentation*. New York: Routledge, 1996.
- Levine, Robert. *Ethics and the Regulation of Clinical Research*. 2nd ed. New Haven, CT: Yale University Press, 1988.
- Macrina, Francis L. *Scientific Integrity: An Introductory Text with Cases*. Washington, DC: American Society for Microbiology Press, 2000.
- Millar, Ronald. *The Piltdown Men*. New York: St. Martin's Press, 1972.
- National Academy of Sciences, National Academy of Engineering, Institute of Medicine. *Advisor, Teacher, Role Model, Friend: On Being a Mentor to Students in Science and Engineering*. Washington, DC: National Academy Press, 1997.
- National Academy of Sciences, National Academy of Engineering, Institute of Medicine. *On Being a Scientist: Responsible Conduct in Research*. Washington, DC: National Academy Press, 1995.
- National Academy of Sciences, National Academy of Engineering, Institute of Medicine. *Responsible Science, Vol. 1: Ensuring the Integrity of the Research Process*. Washington, DC: National Academy Press, 1992.
- “NIH Policy on Instruction in the Responsible Conduct of Research.” National Institutes of Health. Available from [www1.od.nih.gov/](http://www1.od.nih.gov/).
- Nuremberg Code. Available at: [www.hhs.gov/ohrp/references/nurcode.htm](http://www.hhs.gov/ohrp/references/nurcode.htm).
- Penslar, Robin Levin (ed.). *Research Ethics: Cases and Materials*. Bloomington: Indiana University Press, 1995.
- “Requirement for the Instruction in Responsible Conduct of Research in National Research Service Award Training Grants.” *National Institutes of Health Guide* 21 (43) (1992). Available from <http://grants.nih.gov/>.
- Resnik, David B. “Social Epistemology and the Ethics of Research.” *Studies in the History and Philosophy of Science* 27 (1996): 565–86.
- Resnik, David. *The Ethics of Research*. New York: Routledge, 1998.
- Resnik, David B. “From Baltimore to Bell Labs: Reflections on Two Decades of Debate about Scientific Misconduct.” *Accountability in Research* 10 (2003): 123–135.
- Shamoo, Adil E. and David B. Resnik. *Responsible Conduct of Research*. New York: Oxford University Press, 2003.
- Shrader-Frechette, Kristin. *Ethics of Scientific Research*. Lanham, MD: Rowman & Littlefield, 1994.
- Sigma Xi. *Honor in Science*. New Haven, CT: Sigma Xi, The Scientific Research Society, 1986.
- Sigma Xi. *The Responsible Researcher: Paths and Pitfalls*. Research Triangle Park, NC: Sigma Xi, The Scientific Research Society, 1999.
- Singer, Peter. *Animal Liberation: A New Ethics for Our Treatment of Animals*. New York: Random House, 1975.
- Steneck, Nicholas. *ORI Introduction to Responsible Conduct of Research*. Bethesda, MD: Office of Research Integrity, Department of Health and Human Services, 2004.
- “Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication, 2004.” International Committee of Medical Journal Editors. Available from [www.icmje.org/](http://www.icmje.org/).
- Veatch, Robert M. *The Patient as Partner: A Theory of Human Experimentation Ethics*. Bloomington: Indiana University Press, 1987.
- Whitbeck, Caroline. *Ethics in Engineering Practice and Research*. New York: Cambridge University Press, 1998.

David B. Resnik (2005)

## SCIENCE AND PSEUDOSCIENCE

Since the rise of modern science in the sixteenth and seventeenth centuries, attempts to adjudicate the difference between science and pseudoscience have always been more than an exercise in academic debate. The religious, political, and social implications of how science is defined, who defines it, and who and what is left out of the definition has been a contentious one. Today, the term *pseudoscience* is often employed by those in the scientific community to disparage claims to scientific credibility that, in fact, lack evidence or fail to employ the methods of science. *Pseudoscience* is only one term used to contrast with science; others include, on the neutral side, *non-science*, *protoscience*, *prescience*, *frontiers science*, and *borderlands science*; and on the pejorative side, *pathological science*, *junk science*, *voodoo science*, *crackpot science*, and *bad science*.

With the ascendancy of science in the seventeenth century other knowledge traditions began to employ the empirical methods of science to gain respectability. The study of demons, witches, and spirits, for example, took a decidedly empirical turn in the early modern period, out of religious concerns that atheism might ascend to social respectability along with science. One observer wrote, “Atheists abound in these days and witchcraft is called

into question. If neither possession nor witchcraft (contrary to what has been so long generally and confidently affirmed), why should we think that there are devils? If no devils, no God" (Walker, pp. 71–72). By the nineteenth century the study of such quasi-scientific ideas as phrenology, mesmerism, and spiritualism was organized through scientific societies, such as the Society for Psychical Research, founded in London in 1882, and the American Society for Psychical Research, founded in Boston in 1885, both of which included as active members prominent scientists.

### THE DEMARCATION PROBLEM

In the twentieth century the philosophy of science developed into a viable academic discipline, out of which grew attempts to delimit science and nonscience traditions. In *The Logic of Scientific Discovery*, for example, the philosopher of science Karl Raimund Popper identified what he called "the problem of demarcation," that is "the problem of finding a criterion which would enable us to distinguish between the empirical sciences on the one hand, and mathematics and logic as well as 'metaphysical' systems on the other" (1934, p. 27). Most scientists and philosophers use induction as the criterion of demarcation—if one reasons from particular observations or singular statements to universal theories or general conclusions, then one is doing empirical science. Popper's thesis was that induction does not actually provide empirical proof—"no matter how many instances of white swans we may have observed, this does not justify the conclusion that all swans are white" (p. 34)—and that, *de facto*, scientists actually reason deductively, from the universal and general to the singular and particular. But in rejecting induction as the preferred (by others) criterion of demarcation between science and nonscience, Popper was concerned that his emphasis on deduction would lead to an inevitable fuzziness of the boundary line. If a scientific theory can never actually be proven, then is science no different from other knowledge disciplines?

Popper's solution to the problem of demarcation was the criterion of *falsifiability*. Theories are "never empirically verifiable," but if they are falsifiable then they belong in the domain of empirical science. "In other words: I shall not require of a scientific system that it shall be capable of being singled out, once and for all, in a positive sense; but I shall require that its logical form shall be such that it can be singled out, by means of empirical tests, in a negative sense: it must be possible for an empirical scientific system to be refuted by experience" (1934, p. 70).

The theory of evolution, for example, has been accused by creationists as being nonscientific because no one was there to see it happen and biologists cannot observe it in the laboratory because it takes too long. But, in fact, by Popper's criterion of falsifiability, the theory of evolution would be doomed to the trash heap of bad science if, say, human fossil remains turned up in the same geological bedding planes as 300-million-year-old trilobites. No such falsification of evolution has ever been found, and although by Popper's criterion this does not mean that the theory has been proven absolutely, it does mean that it has yet to be falsified, thus placing it firmly in the camp of solid empirical science.

### SCIENCE DEFENDED, SCIENCE DEFINED

The evolution-creationism controversy, in fact, has provided both scientific and legal forms of demarcation between science and pseudoscience. It is one thing for academic scientists and philosophers to debate the definition of science; it is another matter when the U.S. Supreme Court weighs in on the issue. Because evolution could not be excluded from public school science classrooms in the early twentieth century, and because the teaching of religious tenets was deemed unconstitutional in a number of state trials in the middle of the twentieth century, in the latter part of the century creationists began to call their doctrines creation-science. Since academic openness calls for a balanced treatment of competing ideas, they argued, creation-science should be taught side by side with evolution-science. In 1982 creationists succeeded in getting passed the Louisiana Balanced Treatment for Creation-Science and Evolution Science Act. In 1985 the law was struck down in the Federal Court of Louisiana, a decision that was appealed to the U.S. Court of Appeals for the Fifth Circuit. In 1986 the U.S. Supreme Court agreed to hear the case, leading to the publication of a remarkable document that clearly and succinctly adjudicated (literally in this case) the difference between science and pseudoscience.

The document was an *amicus curiae* brief submitted to the court on behalf of seventy-two Nobel laureates in science, seventeen state academies of science, and seven other scientific organizations. The *amicus* brief begins by offering a general definition: "Science is devoted to formulating and testing naturalistic explanations for natural phenomena. It is a process for systematically collecting and recording data about the physical world, then categorizing and studying the collected data in an effort to infer the principles of nature that best explain the observed

phenomena.” Next, the scientific method is discussed, beginning with the collection of “facts,” the data of the world. “The grist for the mill of scientific inquiry is an ever increasing body of observations that give information about underlying ‘facts.’ Facts are the properties of natural phenomena. The scientific method involves the rigorous, methodical testing of principles that might present a naturalistic explanation for those facts” (1986, p. 23).

Based on well-established facts, testable hypotheses are formed. The process of testing “leads scientists to accord a special dignity to those hypotheses that accumulate substantial observational or experimental support.” This “special dignity” is called a “theory” that, when it “explains a large and diverse body of facts” is considered “robust” and if it “consistently predicts new phenomena that are subsequently observed” it is “reliable.” Facts and theories are not to be used interchangeably or in relation to one another as more or less true. Facts are the world’s data. Theories are explanatory ideas about those facts. “An explanatory principle is not to be confused with the data it seeks to explain.” Constructs and other nontestable statements are not a part of science. “An explanatory principle that by its nature cannot be tested is outside the realm of science” (pp. 23–24).

It follows from the nature of scientific method that no explanatory principles in science are final. “Even the most robust and reliable theory ... is tentative. A scientific theory is forever subject to reexamination and—as in the case of Ptolemaic astronomy—may ultimately be rejected after centuries of viability.” Scientists encounter uncertainty as a regular and natural part of their work. “In an ideal world, every science course would include repeated reminders that each theory presented to explain our observations of the universe carries this qualification: ‘as far as we know now, from examining the evidence available to us today’” (1986, p. 24). Science also seeks only naturalistic explanations for phenomena. “Science is not equipped to evaluate supernatural explanations for our observations; without passing judgment on the truth or falsity of supernatural explanations, science leaves their consideration to the domain of religious faith” (p. 23). According to the amicus any body of knowledge accumulated within the guidelines previously described is considered scientific and suitable for public school science education; and any body of knowledge not accumulated within these guidelines is not considered scientific.

On June 19, 1987, the U.S. Supreme Court, by a vote of 7 to 2, held that the Louisiana Act “is facially invalid as violative of the Establishment Clause of the First Amend-

ment, because it lacks a clear secular purpose” and that “[t]he Act impermissibly endorses religion by advancing the religious belief that a supernatural being created humankind” (*Edwards v. Aguillard*, 1987). The Louisiana trial in general, and the amicus brief in particular, had the effect of temporarily galvanizing the scientific community into defining science as a body of knowledge accumulated through a particular scientific method, as defined by the leading members of the scientific community themselves. Science is as scientists do.

## DELIMITING THE BOUNDARIES BETWEEN SCIENCE AND PSEUDOSCIENCE

Creation-science (and its most recent hybrid, intelligent design theory) is just one of many claims that most mainstream scientists reject as pseudoscience. But what about those claims to scientific knowledge that are not so obviously classified as pseudoscience? When encountering a claim, how can one determine whether it constitutes a legitimate assertion as scientific? What follows is a list of ten questions that get to the heart of delimiting the boundaries between science and pseudoscience:

- (1) How reliable is the source of the claim? All scientists make mistakes, but are the mistakes random, as one might expect from a normally reliable source, or are they directed toward supporting the claimants’ preferred beliefs? Scientists’ mistakes tend to be random; pseudoscientists’ mistakes tend to be directional.
- (2) Does this source often make similar claims? Pseudoscientists have a habit of going well beyond the facts, and so when individuals make many extraordinary claims, they may be more than iconoclasts. What one is looking for here is a pattern of fringe thinking that consistently ignores or distorts data.
- (3) Have the claims been verified by another source? Typically, pseudoscientists make statements that are unverified or are verified by a source within their own belief circle. One must ask who is checking the claims and even who is checking the checkers.
- (4) How does the claim fit with what is known about how the world works? An extraordinary claim must be placed in a larger context to see how it fits. When people claim that the pyramids and the Sphinx were built more than 10,000 years ago by an advanced race of humans, they are not pre-

senting any context for that earlier civilization. Where are its works of art, weapons, clothing, tools, and trash?

- (5) Has anyone made an effort to disprove the claim or has only confirmatory evidence been sought? This is the confirmation bias or the tendency to seek confirmatory evidence and reject or ignore disconfirmatory evidence. The confirmation bias is powerful and pervasive. This is why the scientific method, which emphasizes checking and rechecking, verification and replication, and especially attempts to falsify a claim, is critical.
- (6) Does the preponderance of evidence converge on the claimant's conclusion or a different one? The theory of evolution, for example, is proved through a convergence of evidence from a number of independent lines of inquiry. No single fossil or piece of biological or paleontological evidence has the word *evolution* written on it; instead, there is a convergence from tens of thousands of evidentiary bits that adds up to a story of the evolution of life. Creationists conveniently ignore this convergence, focusing instead on trivial anomalies or currently unexplained phenomena in the history of life.
- (7) Is the claimant employing the accepted rules of reason and tools of research or have those rules and tools been abandoned in favor of others that lead to the desired conclusion? UFOlogists exhibit this fallacy in their continued focus on a handful of unexplained atmospheric anomalies and visual misperceptions by eyewitnesses while ignoring that the vast majority of UFO sightings are fully explicable.
- (8) Has the claimant provided a different explanation for the observed phenomena or is it strictly a matter of denying the existing explanation? This is a classic debate strategy: Criticize your opponent and never affirm what you believe to avoid criticism. This strategy is unacceptable in science.
- (9) If the claimant has proffered a new explanation, does it account for as many phenomena as does the old explanation? For a new theory to displace an old theory, it must explain what the old theory did and then some.
- (10) Do the claimants' personal beliefs and biases drive the conclusions or vice versa? All scientists have social, political, and ideological beliefs that potentially could slant their interpretations of the data,

but at some point, usually during the peer-review system, those biases and beliefs are rooted out or the paper or book is rejected for publication.

## THE ENCHANTED GLASS OF SCIENCE

At the dawn of the scientific revolution in the early seventeenth century, the English philosopher Francis Bacon sought to turn away from the scholastic tradition of logic and reason as the sole road to truth, as well as reject the Renaissance quest to restore the perfection of ancient Greek knowledge. In his 1620 work *Novum Organum* (New Tool, contrary to the opinion of Aristotle's *Organon*), Bacon portrayed science as humanity's savior that would inaugurate a restoration of all natural knowledge through a proper blend of observation and logic, data and theory. Bacon understood, however, that there are significant social and psychological barriers that interfere with one's understanding of the natural world, "For the mind of man is far from the nature of a clear and equal glass, wherein the beams of things should reflect according to their true incidence; nay, it is rather like an enchanted glass, full of superstition and imposture, if it be not delivered and reduced" (p. 53). In the end, thought Bacon, science offers the best hope to deliver the mind from such superstition and imposture. Today, science continues to deliver on that hope.

**See also** Evolutionary Theory (Natural Selection); Philosophy of Science, Problems of; Popper, Karl Raimund.

## Bibliography

- Amicus Curiae Brief of 72 Nobel Laureates, 17 State Academics of Science, and 7 Other Scientific Organizations, in Support of Appellees, Submitted to the Supreme Court of the United States, October Term, 1986, as Edwin W. Edwards, in His Official Capacity as Governor of Louisiana, et al., Appellants v. Don Aguillard et al., Appellees, 1986.*
- Bacon, Francis. *Francis Bacon: A Selection of His Works*, edited by Sidney Warhaft. Toronto: Macmillan, 1965.
- Carroll, Robert Todd. *The Skeptic's Dictionary: A Collection of Strange Beliefs, Amusing Deceptions, and Dangerous Delusions*. Hoboken, NJ: Wiley, 2003.
- Charpak, Georges, and Henri Broch. *Debunked!: ESP, Telekinesis, and Other Pseudoscience*. Translated by Bart K. Holland. Baltimore, MD: Johns Hopkins University Press, 2004.
- Cohen, I. Bernard. *Revolution in Science*. Cambridge, MA: Belknap Press of Harvard University Press, 1985.
- Dawkins, Richard. *Unweaving the Rainbow: Science, Delusion, and the Appetite for Wonder*. Boston: Houghton Mifflin, 1998.
- Eddington, Arthur Stanley. *The Philosophy of Physical Science*. Ann Arbor: University of Michigan Press, 1958.

- Feyerabend, Paul. *Against Method: Outline of an Anarchistic Theory of Knowledge*. London: NLB, 1975.
- Feyerabend, Paul. *Science in a Free Society*. London: NLB, 1978.
- Frank, Philipp. *Philosophy of Science: The Link between Science and Philosophy*. Englewood Cliffs, NJ: Prentice-Hall, 1957.
- Gardner, Martin. *Fads and Fallacies in the Name of Science*. New York: Dover Publications, 1957.
- Gardner, Martin. *Science: Good, Bad, and Bogus*. Buffalo, NY: Prometheus Books, 1981.
- Gilovich, Thomas. *How We Know What Isn't So: The Fallibility of Human Reason in Everyday Life*. New York: Free Press, 1991.
- Harré, Rom. *The Philosophies of Science*. 2nd ed. New York: Oxford University Press, 1985.
- Holton, Gerald. *The Advancement of Science and Its Burdens: The Jefferson Lecture and Other Essays*. New York: Cambridge University Press, 1986.
- Kuhn, Thomas S. *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press, 1977.
- Kuhn, Thomas S. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1962.
- Medawar, Peter. *Pluto's Republic*. New York: Oxford University Press, 1984.
- National Science Foundation. "Science Fiction and Pseudoscience." *Science Indicators Biennial Report*, 2002. <http://www.nsf.gov/statistics/seind02/c7/c7h.htm>.
- Nickell, Joe. *The Mystery Chronicles: More Real-Life X-Files*. Lexington: University of Kentucky Press, 2004.
- Olson, Richard. *Science Deified and Science Defied: The Historical Significance of Science in Western Culture*. Vol. 1. Berkeley: University of California Press, 1982.
- Oppenheim, Janet. *The Other World: Spiritualism and Psychical Research in England, 1850–1914*. New York: Cambridge University Press, 1985.
- Park, Robert L. *Voodoo Science: The Road from Foolishness to Fraud*. New York: Oxford University Press, 2000.
- Piattelli-Palmarini, Massimo. *Inevitable Illusions: How Mistakes of Reason Rule Our Minds*. Translated by Massimo, Piattelli-Palmarini, and Keith Botsford. New York: Wiley, 1994.
- Planck, Max. *The Philosophy of Physics*. Translated by H. W. Johnston. New York: Norton, 1936.
- Plait, Philip C. *Bad Astronomy: Misconceptions and Misuses Revealed, from Astrology to the Moon Landing "Hoax"*. New York: Wiley, 2002.
- Popper, Karl. *The Logic of Scientific Discovery*. New York: Harper and Row, 1934.
- Randi, James. *Flim-Flam!: Psychics, ESP, Unicorns, and Other Delusions*. Buffalo, NY: Prometheus Books, 1982.
- Sagan, Carl. *The Demon-Haunted World: Science as a Candle in the Dark*. New York: Random House, 1995.
- Schick, Theodore, Jr., and Lewis Vaughn. *How to Think about Weird Things: Critical Thinking for a New Age*. Mountain View, CA: Mayfield, 1995.
- Shermer, Michael. *The Borderlands of Science: Where Sense Meets Nonsense*. New York: Oxford University Press, 2001.
- Shermer, Michael, ed. *Encyclopedia of Pseudoscience*. Santa Barbara, CA: ABC-CLIO, 2002.
- Shermer, Michael. *Science Friction: Where the Known Meets the Unknown*. New York: Times Books, 2005.
- Shermer, Michael. *Why People Believe Weird Things: Pseudoscience, Superstition, and Other Confusions of Our Time*. New York: W. H. Freeman, 1997.
- Taubes, Gary. *Bad Science: The Short Life and Weird Times of Cold Fusion*. New York: Random House, 1993.
- Vyse, Stuart A. *Believing in Magic: The Psychology of Superstition*. New York: Oxford University Press, 1997.
- Walker, D. P. *Unclean Spirits: Possession and Exorcism in France and England in the Late Sixteenth and Early Seventeenth Centuries*. Philadelphia: University of Pennsylvania Press, 1981.
- Williams, William F., ed. *Encyclopedia of Pseudoscience*. New York: Facts on File, 2000.

Michael Shermer (2005)

## SCIENCE AND RELIGION

See *Religion and the Biological Sciences; Religion and the Physical Sciences*

## SCIENCE POLICY

Science policy deals with how society supports science and how science is utilized in society. The philosophy of science policy considers both interactions from the perspectives of logic, epistemology, ethics, political philosophy, metaphysics, and ontology. Its domain is broader than the philosophy of science, which emphasizes logical and epistemological questions and goes deeper than the descriptive analyses of science, technology, and society (STS) studies.

The central issues in the philosophy of science policy may be distinguished in terms of its two constituent terms: the structure and proper influence of policy on science, and the structure and proper role of science in public policy. Propaedeutic is the question of the nature of policy itself.

### WHAT ARE POLICIES?

What is known as the demarcation problem in the philosophy of science analyzes science as a special form of knowledge. What are known as boundary issues in STS studies describe the distinctive practices of the science-society interface. By contrast, the phenomenon of policy has been subject to little conceptual examination either as knowledge or as practice.

The term *policy* does not occur in traditional political philosophy. There is no word in either Plato's *Republic* or Aristotle's *Politics* that translates as *policy*. Neither

does it occur in Thomas Hobbes's *Leviathan* (1651) or Jean-Jacques Rousseau's *Social Contract* (1762). Indeed, the term is somewhat peculiar to the English language. *Policy* is translated into French as *politique* and into Spanish, depending on context, as *politica* or *norma*. In German it can be rendered by *Politik* and a host of other terms.

In English policies are associated with legal documents such as insurance contracts and guidelines for corporate or governmental behavior. Corporations have policies for the treatment of customers or employees, and governments and government agencies debate military, fiscal, educational, healthcare, and environmental policies. Although a policy has sometimes been defined simply as a decision, this seems inadequate if for no other reason than that one can talk about "policy decisions" and "decision policies."

Reframing Ludwig Wittgenstein's famous question, one may ask what is the difference between my arm going up, me raising my arm (*Philosophical Investigations* §161)—and a policy for raising my arm. The comparison suggests the concept of policy as a guideline for action justified by some kind of analysis. Policies fall in a middle range between decisions about individual actions and general principles for actions. Policies are also to be distinguished from laws and rules.

Since the key difference between my hand going up and me raising my hand is the presence of an intention, a policy might be seen as a particular kind of intention. G. E. M. Anscombe (1957) maintains that for a person to have an intention is to have both a desire to do X and a belief that he or she will do X. On this account intention becomes a secondary rather than a primary phenomenon. In like manner, policies would become secondary phenomena, derivative of desires and beliefs, with the beliefs being justified by scientific evidence or analysis that X will provide results satisfying the desire.

Anscombe's view is criticized by Donald Davidson (1978) and Michael E. Bratman (1987). Davidson believes that intentions are best described as pro-attitudes or evaluative judgments. According to Bratman the most effective way for human beings living in association with others to become effective agents is to have plans, the elements of which are intentions. For Davidson, then, policies might be defined as group pro-attitudes regarding types of actions. For Bratman policies would be closely associated with group plans.

The field of policy studies forms part of a general twentieth-century effort to extend scientific rationality

into group planning, especially in institutional contexts. (Having or making policies applies more to groups of people than to individuals, except for individuals in positions of power who set policies for others.) In this sense all policy is science policy, since it commonly involves scientific justifications of action plans, whether these are for military, fiscal, educational, healthcare, or environmental contexts.

## POLICIES FOR SCIENCE

Following Harvey Brooks (1968), the philosophy of science policy explores two domains: the philosophical aspects of (1) policies for the funding and governance of science, and (2) ways that science can contribute to and/or impede the political process.

For fifty years after World War II, the basic principle underlying U.S. policy for science was that the government should provide no-strings attached funding to scientists, on the grounds that autonomous scientific research invariably benefits society by making contributions to military power, healthcare, and economic competitiveness further down the road (Bush 1945). There were arguments around the margins regarding how much independence to give scientists (e.g., national security required some limits) and about what constituted a well-balanced investment in mathematics, physics, chemistry, biology, and the social sciences. But no debate altered the basic policy: Give money to scientists and let them make their own decisions about how to spend it, because this will eventually rebound to the good of society.

The end of the cold war and increasing budget pressures allowed questions to surface about this basic policy and its foundational justification, the linearity thesis—the belief that autonomous scientific research produces social benefits in an automatic and linear way: more science, more benefit. As historical and sociological analyses of science have shown, however, the linearity thesis applies more to a few highly qualified special cases than as a general rule.

Reassessment of this policy approach has taken multiple forms. In one instance, in response to cases of research misconduct, it has been argued that conscious efforts are needed to promote collaboration between scientists and stakeholders (Guston 2000). Others have asked whether additional knowledge may overwhelm, getting in the way of the reflection needed about alternatives (Mitcham and Frodeman 2002). More generally, STS studies have argued the sociopolitical construction of scientific knowledge, thus challenging the ideal of scientific autonomy.

Taking these reassessments in a political philosophical direction, Philip Kitcher (2001) argues for a modification of linearity policy. Although a moderate realist who sees scientific knowledge as true, Kitcher is not willing to accept existing institutional arrangements for science as the best imaginable. Moreover, given the limitations of public funding, any one scientific research program is necessarily pursued at the expense of others, so that there is a proper place for extrascientific influence on the selection of publicly funded research priorities. Creating the proper policy for science depends on an understanding of what constitutes “well-ordered science” under such conditions.

### SCIENCE IN POLICY MAKING

Several positions have been staked out in terms of how science properly contributes to policy making. In many quarters (both scientific and nonscientific) there has been a strong presumption that science can “answer” policy questions with the definitive account and/or solution to a problem. Although most policy analysts and many scientists now reject any simple version of this belief, it continues to influence the policy-making process. Two basic issues here concern the extent to which science can serve as an assessor or provider of means for nonscientifically determined ends, and whether or not science can assess ends as well as means.

The advancement of external ends has been a vision of modern science since its origins in the work of Francis Bacon, Galileo Galilei, and René Descartes. However, there has been little systematic examination of assumptions about whether in particular cases science is the best way to achieve certain goals. Does increased scientific knowledge or enhanced technologically efficient action always promote social or personal goods? Information overload can, for instance, actually inhibit decision making, and the excitements of technology have been known to skew appreciation of other goods.

At the end of the twentieth century a cadre of scientists and social scientists began to argue that science policy should go beyond the assumption of linearity. Daniel Sarewitz (1996), Donald E. Stokes (1997), and others proposed to examine the publicly stated goals of science funding and then scrutinize whether end-benefit outcomes have been or are likely to be achieved. While this new science policy is a substantial improvement over the old, it nevertheless limps in one important respect: It accepts whatever social goals may have been given a rhetorical blessing by the existing body politic. The philosophical analysis of methods for assessing connections

between scientific effort and assumed end-benefits deserves attention, but it does not reconsider the worthiness of the proposed ends themselves. Ends must be reflected on as well as means—which is where philosophy has a significant role to play.

The most philosophically expansive approach to policy research is what Harold D. Lasswell called the *policy sciences*. In the course of his long, interdisciplinary career, Lasswell sought to develop a method for the systematic analysis of any policy problem (see Lerner and Lasswell 1951, Lasswell 1971). Influenced by the Chicago school pragmatism of such thinkers as George Herbert Mead and Charles E. Merriam, Lasswell’s method centers around five intellectual tasks: clarification of goals; descriptions of trends; analysis of conditions; projection of future developments; and invention, evaluation, and selection of alternatives. These tasks are necessary to address intelligently any number of policy issues, whether public or private, from those associated with taxation or warfare to problems of manufacturing and marketing.

Despite Lasswell’s achievements, however, there are evident opportunities for further philosophical criticism of method in the policy sciences, the practice of which depends on a prior commitment to goods such as human rights and democracy. Illustrations of deeper reflections on ends can be found in such diverse work as Daniel Callahan (2003); Alan Lightman, Daniel Sarewitz, and Christina Desser (2003); and Leon Kass et al. (2003). Callahan questions what he calls the *research imperative* that seems to take every social problem as an opportunity for more scientific research. Lightman, Sarewitz, and Desser undertake a collective reflection on “living with the genie” of scientific and technological productivity. Kass and the President’s Council on Bioethics philosophically assess contemporary aspirations to turn therapy into enhancement.

### PHILOSOPHICAL CRITICISM

In what sense is philosophy of science policy genuine philosophy? Philosophy may be subdivided along two major axes. The first axis is defined by the fundamental questions that constitute philosophical reflection, of which it is common to distinguish at least logic, epistemology, ethics, political philosophy, and metaphysics or ontology. A second axis is constituted by the particular fields or *topoi* where such questions are deployed. This axis yields an indefinite series of regionalizations such as the philosophy of science, of art, of religion, of law, of language, and more. The philosophy of science is characterized by the prominence of logical and epistemological issues, with

only subsidiary attention to ethics, political philosophy, or metaphysics. But, in fact, there are also important questions of the logic of science policy arguments, the character of science policy knowledge, the ethics of science policy decision making, and the political philosophy of science. Because every science policy makes assumptions about the status of science itself, the philosophy of science policy must consider not just the epistemological status of scientific knowledge but also the justice and ontological boundaries of science as a human activity and of its various institutions.

As a regional expression of philosophy, the philosophy of science policy explores a spectrum of concerns. Policy methods deserve logical analysis. The epistemological strengths and weaknesses of models and simulations, not just in physics or climatology but also in policy analysis, call for critical reflection. Behaviors within the professional scientific community and in relations between scientists and the public, including those of policy analysts, require philosophical assessment. Policies in and for scientific communities and those mediating between science and society, along with the role of scientific expertise in a democratic state, are subject to political philosophical scrutiny. Questions related to scientific institutions and their manifold boundary organizations are ontological as well as sociological; the distinction between policy for science and science in policy may be less sound than is commonly assumed.

Discussions that move from interest group power and economic efficiency to questions of truth, goodness, and beauty can make science policy work richer and more robust—and thus, in other than technical or economic senses, more effective. The philosophy of science policy holds out the promise of promoting science policies that are less incomplete, distorted, and unconscious than might otherwise be the case.

**See also** Anscombe, Gertrude Elizabeth Margaret.

### **Bibliography**

- Anscombe, G. E. M. *Intention*. Oxford, U.K.: Blackwell, 1957.
- Bratman, Michael E. *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press, 1987.
- Brooks, Harvey. *The Government of Science*. Cambridge: MIT Press, 1968.
- Bush, Vannevar. *Science: The Endless Frontier*. Washington, DC: U.S. Government Printing Office, 1945.
- Callahan, Daniel. *What Price Better Health? Hazards of the Research Imperative*. Berkeley: University of California Press, 2003.

- Davidson, Donald. "Intending" (1978). In *Essays on Actions and Events*, 83–102. New York: Oxford University Press, 1980.
- Frodeman, Robert, and Carl Mitcham, eds. "Toward a Philosophy of Science Policy: Approaches and Issues." *Philosophy Today*, vol. 48, no. 5 (supplement 2004).
- Guston, David H. *Between Politics and Science: Assuring the Integrity and Productivity of Research*. New York: Cambridge University Press, 2000.
- Kass, Leon R., et al. *Beyond Therapy: Biotechnology and the Pursuit of Happiness*. New York: Regan Books, 2003.
- Kitcher, Philip. *Science, Truth, and Democracy*. New York: Oxford University Press, 2001.
- Lasswell, Harold D. *A Pre-view of Policy Sciences*. New York: American Elsevier, 1971.
- Lerner, Daniel, and Harold D. Lasswell, eds. *The Policy Sciences: Recent Developments in Scope and Method*. Stanford, CA: Stanford University Press, 1951.
- Lightman, Alan, Daniel Sarewitz, and Christina Desser, eds. *Living with the Genie: Essays on Technology and the Quest for Human Mastery*. Washington, DC: Island Press, 2003.
- Mitcham, Carl, and Robert Frodeman. "The Plea for Balance in the Public Funding of Science." *Technology in Society* 24 (1–2) (2002): 83–92.
- Sarewitz, Daniel. *Frontiers of Illusion: Science, Technology, and the Politics of Progress*. Philadelphia: Temple University Press, 1996.
- Stokes, Donald E. *Pasteur's Quadrant: Basic Science and Technological Innovation*. Washington, DC: Brookings Institution Press, 1997.

**Carl Mitcham (2005)**  
**Robert Frodeman (2005)**

## SCIENCE STUDIES

The phrase "science studies" is sometimes used as an umbrella term referring to work in history of science, philosophy of science, research ethics, and so on. But it can also designate a new interdisciplinary approach to the study of science, technology, and society, one that challenges traditional views about the epistemic basis of scientific knowledge and the proper role of science in society. It is this intellectual movement called Science and Technology Studies (STS) that will be discussed here.

Science Studies in the STS sense discards almost all of the distinctions common in traditional philosophy of science, such as the demarcation between the context of discovery and the context of justification, prescriptions versus descriptions, and theory versus observation. Instead, it looks at science as a social activity that cannot be usefully understood in isolation from either technology or society at large. More important than analysis is contextualization. Case studies of local scientific practices are the preferred route to understanding.



What we now call Science Studies (or STS) has a surprising variety of geographical roots. The Strong Program began in Edinburgh in the 1970s. Harry Collins's study of the experimenters' regress was done in England, where Steve Woolgar is also located. Bruno Latour's actor-network theory was developed in Paris. Karin Knorr Cetina worked in Bielefeld, Germany. Feminist standpoint epistemology stems from Sandra Harding's work in America. The 1980s saw a plethora of influential books and articles centered on the notion of social construction, some of it directly influenced by the pioneers such as that of Andrew Pickering, who earned his PhD at Edinburgh, and Trevor Pinch, who worked with Harry Collins at Bath. Systematic criticism of both the historical work and the philosophical claims of STS scholars resulted in the so-called Science Wars, which was triggered by a hard-hitting critical study in 1994 by Paul Gross and Norman Levitt.

Contributors to Science Studies (or STS) draw on resources from diverse disciplinary backgrounds, such as sociology, anthropology, political science, economics, and linguistics. But all the studies share starting points based on interpretations of work in more contemporary philosophy of science. It is to these common philosophical presuppositions that we now turn.

### PHILOSOPHICAL ROOTS

Although Thomas Kuhn himself repudiated some of the most radical extensions of his theory of normal science and scientific revolutions, Science Studies is certainly part of his legacy. There are frequent STS references to his thesis that scientific knowledge is embodied in the practices of a community. Because observation is theory-laden, it is assumed that empirical studies can never be used as a neutral arbiter between rival paradigms. The logical fact that any universal theory is underdetermined by the evidence in support of it is taken to mean that there can be no rational basis sufficient to justify the choice of one theory over another. The point, noted by Pierre Duhem and W.V. Quine, that it is always logically possible to save a theory from refutation by altering auxiliary hypotheses is deployed to advance the skeptical conclusion that philosophical accounts of scientific method and scientific rationality cannot explain why scientists prefer some theories to others and assign great epistemic weight to certain claims and not to others.

Most philosophers of science in the early twenty-first century would agree that there is no ironclad defense against these kinds of skeptical arguments—indeed philosophers are the ones who undermined the quest for

unrevisable foundations in the first place! It is a truism today that there is no instant rationality in science. Instead, philosophers look for the fallible canons of comparative rationality that underlie scientific judgments. For example, in his theory of Scientific Research Programmes, Imre Lakatos argued that most of the history of science could be understood in terms of his methodology. It was only when the evaluations of actual scientists differed from his normative account that one should invoke extra-scientific considerations such as ideology, personal rivalries, or economic interests. Earlier historians of science also drew a distinction between internal history, a narrative of the mostly rational development of scientific ideas through the experimental method, and external history, the story of scientific institutions and their interaction with the larger society. And when sociologists such as Robert Merton studied those scientific institutions, they looked for the operation of special norms that would show why the output of science was superior to the knowledge produced by theologians or grocers!

Adherents to the Science Studies find little use for the internal or external distinction: Trying to isolate ideas from the people who hold them impedes understanding. Similarly, they would invert Lakatos's order of analysis. Instead of looking for the intellectual problems motivating researchers and their attempts to bring evidence to bear on proposed solutions, they begin with a rich sociological description of a scientific episode. What are the lines of authority and collaboration? Through what mechanism is expertise awarded? An important focus of the analysis should be the various rhetorical stratagems employed in communication: How do they reflect the conflicting interests or differing cognitive resources of the participants (for example, theoreticians versus experimenters, policy makers versus scientists, and funding committees versus referees)? The factors that philosophers of science or internalist historians find of interest would be embedded within the STS account, but they would never be privileged. Merton's norms of objectivity and organized skepticism are viewed primarily as part of the rhetorical arsenal of scientists, not as fundamental guides to behavior.

### THE STRONG PROGRAM

Let us now take a brief look at some detailed proposals of the Science Studies approach, bearing in mind that there are disputes within this loosely-knit, interdisciplinary field. We begin with an early, very influential initiative—the Strong Programme of the Edinburgh school. David

Bloor describes the program as a sociology of scientific knowledge (SSK) that is based on four tenets.

First, our accounts of science should be causal (one is reminded of Marxian attempts to explain the content of scientific theories). Secondly, they should be impartial with respect to truth or falsity, rationality or irrationality—both are in need of explanation (many philosophical accounts take rational inference as self-explanatory—in fact to go further may lead to an infinite regress). Furthermore, the explanations of true and false beliefs should be symmetrical; that is, the types of causal factors invoked should be the same (if ideology or political interests are invoked to explain false beliefs, they also need to be brought into a story of the origins of true beliefs). Finally, our approach should be reflexive: The claims made from within the Strong Programme are to be analyzed and explained in exactly the same manner as are episodes in the history of science.

As laid out above, the Strong Programme is a variety of philosophical naturalism, a position that underlies most work in cognitive science. Its most controversial aspect is the symmetry thesis. Why should we posit a priori that the causal chain leading to hallucinations should contain exactly the same elements as the process that produces ordinary visual experiences? Or compare the stories of adherence of Soviet scientists to T. D. Lysenko's theory of acquired characteristics with that of the acceptance of Dmitri Mendeleev's Periodic Table in Czarist Russia: Undoubtedly both situations involved elements of nationalism and the striving of scientists for recognition. But are not the asymmetries more significant? In one case a major factor was the coercion to conform with the wishes of Joseph Stalin; in the other, chemists followed their noses to arrive at a workable classification of chemical phenomena.

Many of the case studies produced by adherents of the Strong Programme focus on scientific controversies. The general pattern is to look at a wide range of social factors, such as class, political pressures, disciplinary commitments, and power structures within the profession. They then argue that these kinds of interests have a strong influence on the conclusions that scientists reach about which of the competing theories is deserving of their allegiance. Their account of the resolution of a controversy does not privilege appeals to epistemic considerations such as predictive accuracy or theoretical coherence.

## SOCIAL CONSTRUCTIONISM

Much work in Science Studies is based on the tenet that all scientific entities are socially constructed. In certain instances, such a claim is nontrivially true: John Searle uses the example of money. A metal disc with Sacagawea's portrait stamped on it does not count as money without the construction of a vast social network that turns it into legal tender. One can also make sense out of the assertion that there were no homosexuals until the late nineteenth century, by adding a gloss to the effect that the term homosexual is to be read as connoting a historically specific, medical-psychological category. But how are we to understand Latour's claim that there was no anthrax before Louis Pasteur's research or that TRH, a product of the hypothalamus, was invented in a certain California laboratory?

It is important to appreciate the difficulty scientists face in isolating natural products, especially when the process involves new sorts of instruments or laboratory procedures. But what is gained by blurring the distinction between a new concept, which certainly is a social construction, a chemical or biological entity, which either existed in nature or was synthesized, and the development of scientific consensus about the match between concept and the object described? It perhaps helps a little bit to understand why Latour and Woolgar would make such perplexing claims if we note that they set out to apply ethnographic approaches to life in the laboratory. Anthropologists who are studying an exotic culture dutifully describe the behavior of the people they are studying. An ethnomethodologist's thick description of a rain dance need say nothing about whether rain actually ensues. In a similar fashion STS scholars can describe the interactions and assertions of scientists without saying anything about whether the object the scientists claim to be studying actually exist. However, radical social constructionists go on to say that social constructions exhaust reality—there is no underlying strata that is being more or less accurately represented.

Even more startling is the so-called Actor-Network Theory (ANT) developed by Latour and Michel Callon, which posits a symmetry between humans and nonhuman entities, such as scallops (Callon's example) or technological devices. These so-called actants form networks in which their competing interests are negotiated. The result is a complex ecological system in which ideas and artifacts, scientist, and resources form an ontology based on what they call relational materiality. Difficult to understand, ANT has generated considerable critical discussion among STS practitioners. Some draw the line at

assigning agency to scallops; others object to attributing stability to inscription devices and the implication that scientific findings involving instruments can be exported from one lab to another without intervening social constructions by the local community.

## FEMINIST STUDIES

A branch of Science Studies that has generated wide interest is the large corpus of feminist writings looking for the effect of gender ideology on both the content and practice of science. These range from significant, but relatively uncontroversial, empirical analyses of the social factors that lead to the attrition of women at every stage of their professional careers to radical claims about the intrinsically sexist nature of the science of mechanics. A central claim, and one that fits in well with some of the STS approaches described above, declares that scientific concepts of sex and gender have historically been strongly influenced by biases inherent in patriarchal societies. One recalls Aristotle's association of the active form with maleness, whereas females were the bearers of passive matter. Anne Fausto-Sterling argues that similar nonscientific influences have entered into the modern study of sex hormones. These are typical examples of social constructionist analyses: Scientific results are held to be strongly influenced by the social milieu; they are not simple reflections of empirical studies.

Feminists posit the influence of gender ideology on the content of science in areas increasingly distant from the study of reproduction. Londa Schiebinger claims that Carl Linneaus's characterization of the class mammalia was influenced by political debates about the propriety of wet nursing. Some have argued that the development of the science of hydrodynamics was delayed because men were uncomfortable dealing with material that was moist and yielding and that the interest in mechanical interactions between hard, rigid bodies that characterized the beginning of the Scientific Revolution was a masculinist preoccupation. And what about the scientific prejudice in favor of linear theories, reductionism, and simplicity? Relying on object-relations theory from psychology, it has even been claimed that the traditional norm of objectivity, of distancing oneself from phenomena, is a reflection of the process by which male children are psychically separated from the mother.

All of these studies follow the STS pattern of trying to show the radical contingency of scientific developments: If social circumstances had been different, the content of science would have been different. Feminists accompany this descriptive analysis with prescriptions for

changing science. Sandra Harding calls for what she calls strong objectivity: If present science is distorted by the predominance of male perspectives, would not science become more objective by the deliberate inclusion of views from the standpoint of women, minorities, workers, and any other group that is underrepresented in today's scientific community? Helen Longino advocates a sort of affirmative action for approaches to understanding the world that are anti-reductionist, nonhierarchical, and unabashedly politically progressive. If science is always socially constructed anyway, why not deliberately construct scientific inquiry in a humanitarian fashion?

## REACTIONS TO SCIENCE STUDIES

The above descriptions of the leading STS approaches give an indication of why their underlying philosophical posits might be viewed as tendentious. Many of their case studies of the factors affecting the acceptance of scientific theories have also generated historiographic skepticism. Critics argue that, contrary to STS claims, Pasteur's religious views, Robert Boyle's preoccupation with chastity, or Karl Pearson's upwardly mobile class interests had a negligible effect on their scientific positions. As an explanation of the acceptance or rejections of scientific hypotheses, STS accounts are not satisfactory.

Ironically, however, the academic reaction to work in Science Studies cannot be understood purely in terms of its intellectual merits—or demerits. Instead, we must also invoke the sorts of interests and ideological factors that STS brings to the forefront! Members of the so-called academic left found the rhetoric of STS very congenial. Already suspicious of the authority of science and troubled by the pace of technological change, they eagerly took up slogans to the effect that science was a creature of the military-industrial complex and a handmaiden to imperialist regimes. The title of a popular textbook by Harry Collins and Trevor Pinch sums it up nicely: *Science as Golem*. There were calls for science for the people, feminist science, and postcolonial science.

Scientists and other intellectuals concerned about the level of funding for scientific research and the general low level of scientific literacy in America mounted a vigorous response. Paul Gross and Norman Levitt's *Higher Superstition: The Academic Left and Its Quarrels with Science* set off intense debates both in the media and in universities. Science Studies practitioners were sometimes lumped in with postmodernists, new age mystics, and so-called scientific creationists. Although STS people sometimes protested that they were not antisience per se, but only objecting to what they considered to be overly adu-

latory accounts of science as a hyperrational activity, they did not make a concerted effort to disassociate themselves from their more radical fellow travelers. Some critics took their ambivalent response to a hoax perpetrated by Alan Sokal, who succeeded in publishing a factually absurd, but politically correct, paper in a leading journal called *Social Text*, as an indication of a weak commitment to traditional scholarly norms.

Science Studies in the STS sense has spurred the attempts of historians, philosophers, and sociologists of science, who favor more traditional approaches to science studies, to provide accounts of the development of science that give us more understanding of the social dimensions of scientific inquiry (see Philip Kitcher's calls for a new socio-historico-philosophical approach). It has also highlighted the importance of developing a more detailed and realistic picture of scientific inquiry. This project is nicely described in the title of Susan Haack's book: *Defending Science—Within Reason: Between Scientism and Cynicism*.

**See also** Feminist Epistemology; Kuhn, Thomas.

### **Bibliography**

- Barnes, Barry, David Bloor, and John Henry. *Scientific Knowledge: A Sociological Analysis*. Chicago: University of Chicago Press, 1996.
- Biagioli, Mario, ed. *The Science Studies Reader*. New York: Routledge, 1998.
- Bloor, David. *Knowledge and Social Imagery*. 2nd ed. Chicago: University of Chicago Press, 1991.
- Callon, Michel. "Some Elements of A Sociology of Translation: Domestication of the Scallops and the Fishermen of St. Brieuç Bay." In *Power, Action, and Belief: A New Sociology of Knowledge?*, edited by J. Law. London: Routledge and Kegan Paul, 1986.
- Collins, Harry M. *Changing Order: Replication and Induction in Scientific Practice*. Chicago: University of Chicago Press, 1992.
- Fausto-Sterling, Anne. *Myths of Gender: Biological Theories about Women and Men*. 2nd ed. New York: Basic Books, 1992.
- Gross, Paul R. and Norman Levitt. *Higher Superstition: The Academic Left and Its Quarrels with Higher Education*. Baltimore, MD: Johns Hopkins Press 1998.
- Haack, Susan. *Defending Science—Within Reason: Between Scientism and Cynicism*. Amherst, NY: Prometheus Books, 2003.
- Hacking, Ian. *The Social Construction of What?* Cambridge, MA: Harvard University Press, 2000.
- Harding, Sandra. *Whose Science? Whose Knowledge? Thinking from Women's Lives*. Ithaca, NY: Cornell University Press, 1991.
- Kitcher, Philip. *Science, Truth, and Democracy*. New York: Oxford University Press, 2003.
- Knorr Cetina, Karin D. *The Manufacture of Knowledge: An Essay on The Constructivist and Contextual Nature of Science*. Oxford: Pergamon Press, 1981.
- Koertge, Noretta, ed. *A House Built on Sand: Exposing Postmodernist Myths about Science*. New York: Oxford University Press, 1998.
- Latour, Bruno. *The Pasteurization of France*. Cambridge, MA: Harvard University Press, 1993.
- Latour, Bruno. *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press, 1988.
- Latour, Bruno, Steve Woolgar, and Jonas Salk. *Laboratory Life*. Princeton, NJ: Princeton University Press, 1986.
- Longino, Helen. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, NJ: Princeton University Press, 1990.
- Merton, Robert K. *The Sociology of Science*. Chicago: University of Chicago Press, 1973.
- Parsons, Keith. *The Science Wars: Debating Scientific Knowledge and Technology*. Amherst, NY: Prometheus, 2003.
- Pickering, Andrew. *Constructing Quarks: A Sociological History of Particle Physics*. Chicago: University of Chicago Press, 1984.
- Pinch, Trevor. *Confronting Nature: The Sociology of Solar-Neutrino Detection*. Dordrecht: Reidel, 1986.
- Pinnick, Cassandra L., Noretta Koertge, and Robert F. Almeder, eds. *Scrutinizing Feminist Epistemology: An Examination of Gender in Science*. New Brunswick, NJ: Rutgers University Press, 2003.
- Potter, Elizabeth. *Gender and Boyle's Law of Gases*. Bloomington: Indiana University Press, 2001.
- Schiebinger, Londa. *Has Feminism Changed Science?* Cambridge, MA: Harvard University Press, 1999.
- Searle, John R. *The Construction of Social Reality*. New York: The Free Press, 1995.
- Seismondo, Sergio. *An Introduction to Science and Technology Studies*. Oxford: Blackwell, 2004.
- Sokal, Alan D., and Jean Bricmont. *Fashionable Nonsense: Postmodern Intellectuals' Abuse of Science*. New York: Picador USA, 1998.

**Noretta Koertge (2005)**

## SCIENTIA MEDIA AND MOLINISM

The *scientia media* is a key term in the theology of Luis de Molina (1535–1600) and in the variants of his teaching introduced by the later Jesuits, especially Robert Bellarmine, Leonard Lessius, Francisco Suárez, and Gabriel Vasquez, in the attempt to resolve the apparent contradiction between the doctrines of grace and of free will.

Molina, a Spanish Jesuit who taught at Coimbra and Evora in Portugal, published his famous *Liberi Arbitrii cum Gratiae Donis, Divina Praescientia, Providentia, Praedestinatione et Reprobatione Concordia* at Lisbon in

1588. The publication of the *Concordia*, as it came to be called, soon led to a controversy that divided the theologians and philosophers of Spain. Generally, the position of Molina was enthusiastically supported by members of his own order and just as vigorously denounced by the Thomists.

For Molina the essential problem was to maintain both human freedom and the efficacy of grace. Given the fact of God's foreknowledge, Molina wished to preserve such a foreknowledge without lapsing into determinism, to show that although God knows infallibly what an individual will freely do, such an infallible knowledge in no way determines the will of the individual. Molina argued that there is a cooperation or concursus of human free will with the divine grace, in contrast to the Thomist view that man's will was physically predetermined to act freely by God. Molina held that this was only a disguised form of determinism. The Thomists maintained that Molina denied the universal divine causality.

The central point in Molina's solution of this problem is based upon the *scientia media*. This, according to Molina, is a form of the divine knowledge that lies between the two forms of God's knowledge that Thomas Aquinas had described in the *Summa*. Thomas maintained that God's knowledge may be one of "vision," a knowledge of that which exists, has existed, or will exist. Alternatively, God's knowledge may consist of the purely possible, a knowledge of "simple understanding," of things and events that have not existed, do not exist, and will not exist. The *scientia media* for Molina is a mean between these two forms of knowledge and is the knowledge that God has of conditional future contingent events; thus, God foreknows from all eternity what an individual would do under certain circumstances if offered his grace. Thomas held that nothing lies outside the divine causality and that God's knowledge, or vision, of the future free acts of the individual entails an act of will by God that predetermines that our acts are free. Molina insisted that God's knowledge is prior to the decree of his will and that his foreknowledge does not predetermine our free acts. God, knowing infallibly what an individual will do under certain circumstances if offered his grace, decrees the circumstances and the grace necessary to effect the cooperative action of the individual. Hence, the infallibility and efficacy of grace is due to the infallibility of God's knowledge, the *scientia media*, not to anything in the grace itself.

The distinction between sufficient and efficacious grace throws further light on these contrasting positions. Like the Thomists, Molina accepted the necessity of grace

for salvation, the absolute gratuity of grace, and that sufficient grace is given to all people. However, Molina denied the need for any distinction between sufficient and efficacious grace. He claimed that sufficient grace becomes efficacious if the will of the individual accepts it. Thus, God foreknew St. Paul's consent before he decreed the grace necessary for conversion. The concurrence of the simultaneous act of the individual and the grace of God replaced the notion that the decree of God is prior to the act of the individual and predetermines it. Thomists objected that this made the efficacy of the divine grace dependent on man rather than on God. Molina declared that the efficacy of grace was unimpaired, for its efficacy or infallibility was extrinsic to the act of the individual and intrinsic in God's foreknowledge. In effect, Molina endeavored to preserve more fully the freedom of the individual without destroying the power of grace; the Thomists were more concerned with preserving the power of grace without destroying the freedom of the individual.

#### LATER MOLINISM

Later Molinism is identified largely with what is termed *congruism*, a theological doctrine reflecting especially the views of Bellarmine and Suárez. Congruism retains the principal features of Molina's theology but modifies it in certain respects. Efficacious grace is equated with *gratia congrua* and sufficient grace with *gratia incongrua*. This distinction emphasized more strongly that grace was efficacious when it was congruous with those circumstances and the disposition of the individual that would enable him to will a certain act freely but infallibly. Grace was inefficacious when it was not congruous with the circumstances and disposition of the individual. The efficacy of the *gratia congrua* is intrinsic to the *scientia media* and extrinsic to the will of the individual. *Gratia incongrua* is grace that is sufficient for a salutary act but which the individual will reject.

On predestination the Molinists agreed with the Thomists that God wishes all people to be saved and that he extends sufficient grace to all, that contrary to Pelagianism predestination is wholly gratuitous, and that some individuals are elected in preference to others solely as God wills. However, they tended to modify the Thomist view of an absolute predestination to glory irrespective of foreseen merits. Many of the Molinists argued that predestination is conditional upon the future actions of the individual and becomes absolute only with the foreseen merits of the individual. In contrast to the Thomists, who argued for the priority of predestination

to glory to the predestination of efficacious grace, the Molinists held that God foresees in the *scientia media* that some will cooperate with his grace and predestined them to glory by offering them his grace.

The differences between Molina and his successors are more often subtle than essential. Although the debate on Molinism has continued for more than three centuries, Molinism is clearly compatible with faith and continues to have many supporters. Like Thomism it has its difficulties and its critics. The difference between the two schools remains essentially one of the relative emphasis to be placed upon grace or freedom.

**See also** Bellarmine, St. Robert; Molina, Luis de; Suárez, Francisco; Vasquez, Gabriel.

### Bibliography

- Beilby, James K., and Paul R. Eddy. *Divine Foreknowledge: Four Views*. Downers Grove, IL: InterVarsity Press, 2001.
- Bonet, A. *La filosofía de la libertad en los controversias teológicas del siglo XVI. y primera mitad del XVII*. Barcelona, 1932.
- Boyer, C. "Providence et liberté." *Gregorianum* (1938): 194–208.
- Craig, William Lane. "No Other Name: A Middle Knowledge Perspective on the Exclusivity of Salvation Through Christ." In *The Philosophical Challenge of Religious Diversity*, edited by Philip L. Quinn and Kevin Meeker. New York: Oxford University Press, 2000.
- Craig, William Lane. "Creation, Providence and Miracles." In *Philosophy of Religion*, edited by Brian Davies. Georgetown University Press: Washington, 1998.
- Dekker, Eef. *Middle Knowledge*. Leuven: Peeters, 2000.
- Flint, Thomas P. *Divine Providence: The Molinist Account*. Ithaca: Cornell University Press, 1998.
- Flint, Thomas. "Middle Knowledge and the Doctrine of Infallibility." In *Philosophical Perspectives, 5: Philosophy of Religion*, edited by James E. Tomberlin. Ridgeview: Atascadero, 1991.
- Flint, Thomas P. "The Multiple Muddles of Maverick Molinism." *Faith and Philosophy* 20 (1) (2003): 91–100.
- Flint, Thomas P. "The Possibilities of Incarnation: Some Radical Molinist Suggestions." *Religious Studies* 37 (3) (2001): 307–320.
- Hasker, William. "The Antinomies of Divine Providence." *Philosophia-Christi* 4 (2) (2002): 361–375.
- Hasker, William. "A New Anti-Molinist Argument." *Religious Studies* 35 (3) (1999): 291–297.
- Hasker, William. "Providence and Evil: Three Theories." *Religious Studies* 28 (1) (1992): 91–105.
- Kleinhapp, Johann. *Der Staat bei Ludwig Molina*. Innsbruck: Rauch, 1935.
- Linville, Mark. "Ockhamists and Molinists in Search of a Way Out." *Religious Studies* 31 (4) (1995): 501–515.
- Myers, David B. "Exclusivism, Eternal Damnation, and the Problem of Evil: A Critique of Craig's Molinist Soteriological Theodicy." *Religious Studies* 39 (4) (2003): 407–419.
- O'Connor, Timoth. "The Impossibility of Middle Knowledge." *Philosophical Studies* 66 (2) (1992): 139–166.
- Pegis, Anton C. "Molina and Human Liberty." In *Jesuit Thinkers of the Renaissance*, 75–132. Milwaukee: Marquette University Press, 1939.
- Perszyk, Kenneth J. "An Anti-Molinist Argument." *Philosophical Studies* 90 (3) (1998): 215–235.
- Perszyk, Kenneth J. "Free Will Defence With and Without Molinism." *International Journal for Philosophy of Religion* 43 (1) (1998): 29–64.
- Perszyk, Kenneth J. "Molinism and Compatibilism." *International Journal for Philosophy of Religion* 48 (1) (2000): 11–33.
- Perszyk, Kenneth J. "Molinism and the Consequence Argument: A Challenge." *Faith and Philosophy* 20 (2) (2003): 131–151.
- Perszyk, Kenneth J. "Molinism and Theodicy." *International Journal for Philosophy of Religion* 44 (3) (1998): 163–184.
- Rasolo, Louis. "Le dilemme du concours divin." *Analecta Gregoriana* 80 (1956).
- Romeyer, B. "Libre arbitre et concours selon Molina." *Gregorianum* (1942): 169–201.
- Stegmüller, Francis, ed. *Geschichte des Molinismus*, I: Neue Molinaschriften. Münster: Aschendorff, 1935.
- Vansteenbergh, E. "Molinisme." In *Dictionnaire de Théologie Catholique*, Vol. X, cols. 2184–2187. Paris, 1929.
- Werther, David. "Open Theism and Middle Knowledge: An Appraisal of Gregory Boyd's Neo-Molinism." *Philosophia Christi* 5 (1) (2003): 205–215.
- Zimmerman, Dean. "Richard Gale and the Free Will Defense." *Philo* 6 (1) (2003): 78–113.

John A. Mourant (1967)

Bibliography updated by Tamra Frei (2005)

## SCIENTIFIC EXPLANATION

See *Explanation*

## SCIENTIFIC LAWS

See *Laws, Scientific*

## SCIENTIFIC METHOD

What follows is a description of various views on inductive inference and methods for inferring general theories as they have developed from the scientific revolution to modern times. Later, the development of methods for discovering causal relationships will be discussed.

**MODERN METHODOLOGY.** A strong influence on contemporary methodology is interdisciplinary research. In

the twentieth century, the question of how we can use observations to attain empirical knowledge became the subject of research in a number of disciplines, such as statistics, econometrics, and computer science. Modern philosophy of method continues to contribute to and draw on developments in related disciplines.

Another strong influence on contemporary methodology arises from studies of the history of science, which captured the attention of philosophers because of the groundbreaking work of Thomas Kuhn (1922–1996) on the *Structure of Scientific Revolutions*. Kuhn argued that scientific textbook accounts of the history of science as a wholly progressive series of discoveries are false for scientific revolutions. His work has suggested that changes of method across revolutions undercut attempts to apply common standards to evaluate prerevolution and postrevolution theories.

Kuhn also criticized the methodological ideas of Karl Popper (1902–1994). Popper had asked the question of what distinguishes (“demarcates”) scientific hypotheses from nonscientific hypotheses. He emphasized that science proceeds by testing hypotheses against empirical data, and thus located the characteristic of scientific hypotheses in their empirical testability. Popper’s basic view of testing a hypothesis against data was to derive predictions from the hypothesis and see if they matched the data (conjectures and refutations). If the data does not match the predictions, they falsify the hypothesis.

This led Popper to postulate that scientific hypotheses must be falsifiable. Popper’s falsifiability criterion has been very influential, arguably more outside of the philosophy of science than inside. Kuhn objected to the falsifiability concept because, according to him, history shows that scientists do not subject major scientific theories (or paradigms) to falsification. Instead, scientists view a mismatch between theory and data as an anomaly, a puzzle to be resolved by further research. Many philosophers of science took Kuhn’s moral to be that logic-based analyses of scientific method cannot capture the dynamics of major scientific change. Scientific revolutions would instead be determined by complex sociopolitical processes within the scientific community, played out within the specific historical context. Modern methodologists aim to avoid both the extremes of a context-free universal scientific logic on the one hand, and an entirely context-specific study of particular historical episodes on the other.

## METHOD IN THE SCIENTIFIC REVOLUTION

Two topics of inquiry held center stage during the scientific revolution: the traditional problems of astronomy, and the study of gravity as experienced by bodies in free fall near the surface of the earth. Johannes Kepler (1571–1630) proposed that the predictive empirical equivalence between geocentric and heliocentric world systems that holds in principle could be offset by appeal to physical causes (Jardine 1984). He endorsed the appeal by Nicolas Copernicus (1473–1543) to the advantage offered his system from agreeing measurements of parameters of the earth’s orbit from several retrograde motion phenomena of the other planets (1596/1981). In his classic marshaling of fit to the impressive body of naked eye instrument observation data by Tycho Brahe (1546–1601), Kepler appealed to this advantage as well as qualitative intuitions about plausible causal stories and intuitions about cosmic harmony to arrive at his ellipse and area rules (1609/1992). He later arrived at his harmonic rule (1619/1997). His Rudolphine Tables of 1627 were soon known to be far more accurate than any previously available astronomical tables (Wilson 1989).

Galileo Galilei (1564–1642) described his discovery of Jupiter’s moons and exciting new information about our moon in the celebrated report of his telescope observations (1610/1989). His later observations of phases of Venus provided direct observational evidence against Ptolemy’s system, though not against Tycho’s geoheliocentric system. This was included in his argument for a Copernican heliocentric system in his famously controversial *Dialogue* (1632/1967).

Galileo’s study of gravity faced the challenge that because of complicating factors such as air resistance one could not expect the kind of precise agreement with measurement that was available in astronomy. In his celebrated *Two New Sciences* (1638/1914), Galileo proposed uniformly accelerated fall as an exact account of idealized motion that would obtain in the absence of any resistant medium, even though the idealization is impossible to actually implement. He argues that the perturbing effects of resistance are too complex to be captured by any theory, but that the considerations he offers, including inclined plane experiments that minimize the effects of resistance, support his idealized uniformly accelerated motion as the principal mechanism of such terrestrial motion phenomena as free fall and projectile motion.

An important part of what distinguishes what we now characterize as the natural sciences is the method exemplified in the successful application of universal

gravity to the solar system. Isaac Newton (1642–1727) characterizes his laws of motion as accepted by mathematicians and confirmed by experiments of many kinds. He appeals to propositions inferred from them as resources to make motion phenomena measure centripetal forces. These give systematic dependencies that make the areal law for an orbit measure the centripetal direction of the force maintaining a body in that orbit, that make the harmonic law for a system of orbits about a common center, and that make the absence of orbital precession (not accounted for by perturbations) for any such orbit, measure the inverse square power for the centripetal force. His inferences to inverse-square forces toward Jupiter, Saturn, and the sun from orbits about them are inferences to inverse-square centripetal acceleration fields backed up by such measurements.

Newton's moon-test shows that the length of a seconds pendulum at the surface of the earth and the centripetal acceleration of the moon's orbit count as agreeing measurements of a single earth-centered inverse-square acceleration field. On this basis Newton identified the force maintaining the moon in orbit with terrestrial gravity. His first two rules endorse this inference. Rule number one states "no more causes of natural things should be admitted than are both true and sufficient to explain their phenomena" (Newton 1726/1999, p. 794). Rule number two adds that, therefore, "the causes assigned to natural effects of the same kind must be, so far as possible, the same" (Newton 1726/1999, p. 795).

Newton argues that all bodies gravitate toward each planet with weights proportional to their masses. He adduces a number of phenomena that give agreeing measurements of the equality of the ratios of weight to mass for bodies at equal distances from planets. These include terrestrial pendulum experiments and the moon-test for gravitation toward the earth, as well as the harmonic laws for orbits about them for gravitation toward Saturn, Jupiter, and the sun. They also include the agreement between the accelerations of Jupiter and its satellites toward the sun, as well as between those of Saturn and its satellites and those of the earth and its moon toward the sun.

His third rule endorses the inference that these all count as phenomena giving agreeing measurements of the equality of the ratios of weight to mass for all bodies at any equal distances from any planet whatsoever. Rule number three states that "those qualities of bodies that cannot be intended and remitted (i.e., qualities that cannot be increased and diminished) and that belong to all bodies on which experiments can be made should be

taken as qualities of all bodies universally" (Newton 1726/1999, p. 795).

Newton's fourth rule added that "In experimental philosophy propositions gathered from phenomena by induction should be considered either exactly or very nearly true notwithstanding any contrary hypothesis until yet other phenomena make such propositions either more exact or liable to exceptions" (Newton 1726/1999, p. 796). This rule was added to justify treating universal gravity as an established scientific fact, notwithstanding complaints that it was unintelligible in the absence of a causal explanation of how it results from mechanical action by contact.

Newton's inferences from phenomena exemplify an ideal of empirical success as convergent accurate measurement of a theory's parameters by the phenomena to be explained. In rule four, a mere hypothesis is an alternative that does not realize this ideal of empirical success sufficiently to count as a serious rival.

Rule four endorses provisional acceptance. Deviations count as higher order phenomena carrying information to be exploited. This method of successive corrections guided by theory mediated measurement led to increasingly precise specifications of solar system phenomena backed up by increasingly precise measurements of the masses of the interacting solar system bodies.

This notion of empirical success as accurate convergent theory mediated measurement of parameters by empirical phenomena clearly favors the theory of general relativity of Albert Einstein (1879–1955) over Newton's theory (Harper 1997). Moreover, the development and application of testing frameworks for general relativity are clear examples of successful scientific practice that continues to be guided by Newton's methodology (Harper 1997, Will 1986 and 1993). More recent data such as that provided by radar ranging to planets and lunar laser ranging provide increasingly precise post Newtonian corrections that have continued to increase the advantage over Newton's theory that Newton's methodology would assign to general relativity (Will 1993).

## HYPOTHETICO-DEDUCTIVISM

In the preface to his *Treatise on Light*, Christian Huygens (1629–1695) provided a nice characterization of the hypothetico-deductive (H-D) alternative to Newton's method:

There will be seen in it demonstrations of those kinds which do not produce as great a certitude



as those of Geometry, and which even differ very much therefrom, since whereas the Geometers prove their Propositions by fixed and incontestable Principles, here the Principles are verified by the conclusions to be drawn from them; the nature of these things not allowing of this being done otherwise. It is always possible to attain thereby to a degree of probability which very often is scarcely less than complete proof. To wit, when those things which have been demonstrated by the Principles that have been assumed correspond perfectly to the phenomena which experiment has brought under observation; especially when there are a great number of them, and further, principally, when one can imagine and foresee new phenomena which ought to follow from the hypotheses which one employs, and when one finds that therein the fact corresponds to our prevision.

(HUYGENS 1690/1962, P. VI AND VII)

Thus H-D method construes empirical success as success in prediction. The limitation of empirical success to prediction alone has suggested to some philosophers of science that distinguishing between theories that agree on predictions would have to be based on nonempirical criteria.

### PREDICTED FIT TO FUTURE DATA

Given plausible assumptions about errors in data, a model that fits a given body of data too closely is likely to be tracking random errors in the data in addition to the lawlike phenomenon under investigation. Statisticians refer to this as “overfitting the data.” They have designed many criteria to reveal cases where a simpler model has better expected fit-to-future data generated by repetitions of an experiment than a more complex model that better fits the data so far. Among philosophers of science, Malcolm Forster and Elliott Sober have appealed to the Akaike Information Criterion to challenge the assumption that fit-to-past data exhausts the criteria for scientific inference. This criterion is not sufficient to recover Newton’s method (Myrvold and Harper 2002). The extent to which other such proposals can recover Newton’s method is an open question.

### BAYESIAN METHODS

Central to the Bayesian methods is epistemic probability, a rational agent’s degree of belief. A number of arguments have been put forward to defend the probability axioms

as coherence conditions for rational degrees of belief, in analogy to the way logical consistency can be taken as a coherence condition for rational acceptance. Dutch book arguments have shown that degrees of belief violating the probability axioms would assign positive expectations to each bet in a system of bets and conditional bets that would result in sure loss if they were all made together. A number of other arguments for this synchronic condition on rational degrees of belief have been advanced (particularly by Frank Plumpton Ramsey, Leonard J. Savage, Abner Shimony, Bas van Fraassen, Richard T. Cox, Irving John Good, and J. Aczel).

David Lewis (1941–2001) provided a diachronic Dutch book argument (published in Teller 1976) to defend the Bayesian conditionalization learning model, according to which assigning new degrees of belief given by  $P'(B) = P(B \& A) / P(A)$  is the appropriate response to a learning experience in which the total relevant empirical input is to accept A as new evidence. In 1984 van Fraassen (1941–) extended this diachronic Dutch book argument to defend a condition he called reflection. His proposal to treat the reflection condition as a constraint on degrees of belief that could be counted as rational has led to much controversy.

One central Bayesian theme has been to investigate conditions under which evidence leads to convergence of opinion. Bruno de Finetti (1906–1985) specified conditions that would lead Bayesian agents, who update by repeated conditionization on the outcomes of the same observations, to converge toward agreement in their degrees of belief, however otherwise divergent their prior degrees of belief may have been (1937/1980). Brian Skyrms (1990) has given what is probably the most general possible version of de Finetti’s condition for convergence.

In 2003 Wayne Myrvold (1963–) argued that, for Bayesians, the degree to which a hypothesis unifies phenomena contributes to the degree to which these phenomena support the hypothesis. This suggests that Bayesians can recover important aspects of Newton’s method. It may well be that investigating the representation of Newton’s method of provisional acceptance in a Bayesian model will result in enriching the Bayesian framework to make it offer more resources for illuminating scientific method.

### CAUSATION, CORRELATION, EXPERIMENTATION

In his famous methods (1843), John Stuart Mill (1806–1873) combined ideas about causal inference previously

proposed by John Duns Scotus (1265/66–1308), William Ockham (1280–1349) and Francis Bacon (1561–1626). The work of twentieth century statisticians such as Jerzy Neyman (1894–1981), Karl Pearson (1857–1936), and Ronald A. Fisher (1890–1962) addressed two major shortcomings of Mill's method.

First, Mill assumed that we would observe deterministic causal relationships: Given the cause, the effect must follow every time. However, in a complex situation we typically do not have a complete specification of all operative causes, so we expect to observe trends rather than necessary relationships. For example, although smoking causes lung cancer, it does not do so in every person, because people's physiology varies. Rather, what we observe is a strong association between smoking and lung cancer: Among smokers, the incidence of lung cancer is much higher than among nonsmokers. To define precisely the intuitive notion of "strong association," statisticians developed the concept of correlation, which defines degrees of association (DeGroot 1975).

A second deficiency in Mill's methods is that they fail in the presence of common causes (*confounders* in statistical terminology). For example, suppose we observe that children who play violent video games are more prone to aggressive behavior than children who do not. Mill's logic would lead us to infer that playing violent video games causes aggressive behavior. But another possibility is that the correlation is because of personality traits: that children with an aggressive nature are drawn to violent video games and tend toward aggressive behavior; a preference for violent video games does not cause the behavior, but is merely a symptom of preexisting aggressive tendencies. If this alternative explanation is true, then Mill's methods lead us to the wrong conclusion. The policy implications are significant: If there is a direct causal relationship between video games and aggressive behavior, we expect to reduce aggressive behavior by restricting the availability of video games. But if personality is the underlying common cause of both, restricting access to video games should not decrease aggressive behavior.

A great advance for the problem of unobserved common causes was Fisher's revolutionary idea of the randomized experiment. Suppose that we have the ability to randomly assign half of a group of children to playing violent video games (the treatment group) and the other half to playing something else (the control group). For example, we might flip a coin for each participating child to make the assignment. Then we expect that personality traits, such as a tendency to aggression, would be randomly distributed in each half so that the children play-

ing the video games would, on average, have no more aggressive personalities than the children playing something else. Under those circumstances, if we still find that significantly more of the video game players engage in aggressive behavior than the children playing something else, we can infer a direct causal relationship.

The idea of using randomization to rule out unobserved common causes has been applied in countless practical problems of causal inference, from clinical studies of the effectiveness of medical treatments to experiments for agricultural methods. It has been a most effective tool for addressing the problem of unobserved common causes that besets many of the traditional philosophical proposals for causal inference.

The power of randomization is available only when we have the ability to experimentally create the conditions we wish to investigate. In many settings of interest, we cannot perform experiments but can only passively gather data (these are called "observational studies" in statistics). A prominent physical science based on passive observation is astronomy. Many examples occur in the social sciences and economics. For instance, an economist cannot randomly assign inflation rates to various countries to study how inflation affects employment. A recent set of examples comes from computer science: While many companies gather vast amounts of data about their customers and the transactions they engage in, they rarely have the ability to assign customers randomly to various conditions (e.g., household income).

Philosophers continued to refine their understanding of the relationship between correlation and causation in nonexperimental settings. The work of Hans Reichenbach (1891–1953), published in 1956, was seminal. Reichenbach expounded the common cause principle: roughly, for every correlation between two events A and B, there is some causal explanation that posits either that one is a cause of the other (e.g., A causes B) or that A and B share a common cause. Reichenbach argued that the assumption that significant associations or correlations have causal explanations is deeply ingrained in our scientific and everyday reasoning. Another important concept of Reichenbach was the notion of screening off. The purpose of this concept is to capture the distinction between immediate and intermediate causes in terms of correlations.

For example, suppose that tar content in lungs is the direct cause of cancer, while smoking directly causes tar to accumulate in the lungs, and thereby indirectly causes lung cancer. Then we would observe a correlation between smoking and lung cancer; but knowing the tar

content of the lung would make smoking irrelevant to lung cancer. By contrast, even if we knew whether a subject smokes, the tar content of one's lungs would still be relevant to, or correlated with, the subject getting lung cancer. In Reichenbach's terms, information about tar content screens off information about smoking from conclusions about lung cancer. Because tar content screens off smoking from lung cancer, but not vice versa, Reichenbach suggested that such evidence rules out smoking as a direct cause of lung cancer, and allows us to infer that the effects of smoking are mediated through tar in the lungs.

The philosophers of science—Peter Spirtes, Clark Glymour, and Richard Scheines—developed Reichenbach's ideas about the relationships between correlation and causation using the framework of causal graphs or diagrams (Spirtes 1993). A causal graph is an intuitive representation of causal relationships, in which direct causes are connected with their effects by arrows pointing from cause to effect.

Using the language of causal graphs, Spirtes, Glymour, and Scheines gave a precise formulation of Reichenbach's precept that direct causes screen off indirect ones, known as the Markov condition (I-map in computer science terminology). The common cause principle—that there is no correlation without causation—can be formulated as another principle about diagrams, termed faithfulness (perfect I-map in computer science terminology). Given these principles relating causation and correlation, it is possible to characterize when valid inferences about causal relationships can be drawn from passive observation of associations. The theory is powerful and precise enough to develop computer programs that perform these inferences automatically (the TETRAD system, for instance). With such a program, we can analyze the kind of large datasets that we find in practice, realizing the vision of Bacon and Mill of applying causal inference methods to extensive observation histories.

In computer science, causal diagrams (often called Bayes Nets) have been firmly established as a scheme to capture and reason about associations and causal relationships, giving rise to thriving commercial developments with many practical applications (Pearl 1988, 2000). Econometrics, the study of statistical methods for economic problems, has a rich tradition of developing methods for nonexperimental causal inference going back to the early twentieth century (path diagrams and structural equation models). It turns out that many of these ideas and techniques can be seen as instances of

causal diagram methods (Pearl 2000). While the theory of causal inference from passive observation is not yet as firmly established as the methodology based on randomization, at the beginning of the twenty-first century we see a common framework emerging shared and sustained by philosophy, computer science, and economics.

**See also** Bayes, Bayes' Theorem, Bayesian Approach to Philosophy of Science; Philosophy of Statistical Mechanics; Scientific Revolutions.

### Bibliography

- Aczel, J. *Lectures on Functional Equations and Their Applications*. New York: Academic Press, 1966.
- Cox, R. *The Algebra of Probable Inference*. Baltimore, MD: John Hopkins Press, 1961.
- de Finetti, B. "Foresight: Its Logical Laws, Its Subjective Sources" (1937). Translated by H. E. Kyburg and H. Smokler. In *Studies in Subjective Probability*. Huntington, NY: Kreiger, 1980.
- DeGroot, Morris H. *Probability and Statistics*. Reading, MA: Addison-Wesley, 1975.
- Earman, J. *Bayes or Bust?: A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: MIT press, 1992.
- Forster, M., and E. Sober. "How to Tell When Simpler, More Unified, or Less Ad Hoc Theories Will Provide More Accurate Predictions." In *British Journal for The Philosophy of Science* 45 (1994): 1–35.
- Galilei, G. *The Sidereal Messenger* (1610). Translated by A. Van Helden. Chicago: University of Chicago Press, 1989.
- Galilei, G. *Dialogue concerning the Two Chief World Systems* (1632). Translated by Stillman Drake. Los Angeles: University of California Press, 1967.
- Galilei, G. *Two New Sciences*. Translated by H. Crew and A. De Salvio. New York: Dover, 1914.
- Good, I. J. *Probability and The Weighing of Evidence*. London: Griffin, 1950.
- Harper, W. L. "Isaac Newton on Empirical Success and Scientific Method." In *The Cosmos of Science: Essays of Exploration*, edited by J. Earman and J. D. Norton. Pittsburgh, PA: University of Pittsburgh Press, 1997.
- Harper, W. L., and C. A. Hooker, eds. *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*. Vol. 1. Dordrecht; Boston: D. Reidel, 1976.
- Huygens, C. *Treatise On Light* (1690). Translated by S. P. Thompson. New York: Dover, 1962.
- Jardine, N. *The Birth of History and Philosophy of Science: Kepler's "A Defense of Tycho against Ursis"*. Cambridge, MA: Cambridge University Press, 1984.
- Jeffreys, H. *Scientific Inference*. 3rd ed. Cambridge, MA: Cambridge University Press, 1973.
- Kepler, J. *The Harmony of The World* (1619). Translated by E. J. Aiton, A. M. Duncan, and J. V. Field. Philadelphia: American Philosophical Society, 1997.
- Kepler, J. *New Astronomy* (1609). Translated by W. H. Donahue. New York: Cambridge University Press, 1992.

- Kepler, J. *The Secret of the Universe* (1596). Translated by A.M. Duncan. New York: Abaris Books, 1981.
- Kuhn, T. S. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1962.
- Kyburg, H. E. *Science & Reason*. New York: Oxford University Press, 1990.
- Myrvold, W. C. "A Bayesian Account of The Virtue of Unification." *Philosophy of Science* 70 (2003): 399–423.
- Myrvold, W. C., and W. Harper. "Model Selection, Simplicity, and Scientific Inference." *Philosophy of Science* 69 (2002): 135–149.
- Newton, I. *Mathematical Principles of Natural Philosophy*. 3rd ed. (1726). Translated by I. B. Cohen and A. Whitman. Los Angeles: University of California Press, 1999.
- Pearl, J. *Causality*. San Mateo, CA: Morgan Kaufmann, 2000.
- Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1988.
- Popper, K. R. *Conjectures and Refutations: The Growth of Scientific Knowledge* (1962). New York: Harper & Row, 1963.
- Popper, K. R. *The Logic of Scientific Discovery* (1959). New York: Harper & Row, 1965.
- Ramsey, F. P. "Truth and Probability." In *The Foundations of Mathematics and Other Logical Essays*. London: Kegan Paul, 1931.
- Reichenbach, H. *The Direction of Time*. Berkeley, CA: University of Los Angeles Press, 1956.
- Reichenbach, H. *The Theory of Probability*. London: Cambridge University Press, 1949.
- Savage, L. J. *Foundations of Statistics*. New York: Wiley, 1954.
- Shimony, A. *Search for a Naturalistic World View: Scientific Method and Epistemology*. Vol. 1. Cambridge, U.K.: Cambridge University Press, 1993.
- Skyrms, B. *The Dynamics of Rational Deliberation*. Cambridge, MA: Harvard, 1990.
- Spiro, P., C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. New York: Springer Verlag, 1993.
- Teller, O. "Conditionalization, Observation, and Change of Preference." In *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, edited by W. L. Harper and C. A. Hooker. Dordrecht; Boston: D. Reidel 1 (1976): 205–257
- van Fraassen, B. "Belief and The Will." *Journal of Philosophy* 81 (1984): 235–256.
- van Fraassen, B. "Calibration: A Frequency Justification for Personal Probability." In *Physics, Philosophy, and Psychoanalysis: Essays in Honor of Adolf Grünbaum*, edited by R. S. Cohen and L. Laudan. Dordrecht: Reidel, 1983.
- Will, C. M. *Theory and Experiment in Gravitational Physics*. 2nd ed. Cambridge, U.K.: Cambridge University Press, 1993.
- Will, C. M. *Was Einstein Right? Putting General Relativity to the Test*. New York: Basic Books, 1986.
- Wilson, C. "Predictive Astronomy in The Century after Kepler." In *The General History of Astronomy: Planetary Astronomy from the Renaissance to the Rise of Astrophysics, Part A: Tycho Brahe to Newton*. Vol. 2, edited by R. Taton and C. Wilson. Cambridge, U.K.: Cambridge University Press, 1989.

*William Harper (2005)*  
*Oliver Schulte (2005)*

## SCIENTIFIC REALISM

Scientific realism is a philosophical view about science that consists of three theses:

*The metaphysical thesis:* The world has a definite and mind-independent structure.

*The semantic thesis:* Scientific theories should be taken at face value. They are truth-conditioned descriptions of their intended domain, both observable and unobservable. Hence, they are capable of being true or false. The theoretical terms featured in theories have putative factual reference.

*The epistemic thesis:* Mature and predictively successful scientific theories are well confirmed and (approximately) true of the world. So the entities posited by them, or entities very similar to those posited, inhabit the world.

## METAPHYSICS

Let us call the first thesis of scientific realism *metaphysical realism*. What exactly is involved in the claim of mind-independence? One way to construe the opposite claim that the world is mind-dependent, along the lines of traditional idealism and phenomenalism, is to argue that the world consists of mental entities, be they ideas or actual and possible sense-data. Thus understood, mind-dependence is a thesis about the kind of stuff that makes up the world. The insistence of scientific realism on metaphysical realism might be thought of as opposing this idealist or phenomenalist doctrine. It might be seen as a declaration that there is nonmental stuff in the world and, in particular, that the entities posited by scientific theories are material. This view is certainly part of the realist construal of mind-independence, but there is more.

There is another, more complicated and interesting, way to construe the claim that the world is mind-dependent. This way centers not on what *types* of entity exist (whether they are material or mental or what have you) but rather on what is involved in claiming that they *exist*. There is a long antirealist philosophical tradition according to which it does not make sense to assert the existence (or reality) of some entities unless we understand this assertion to mean that . . . , where the ellipsis is filled with a suitable epistemic/conceptual condition. Much like realism, these views (call them varieties of verificationist antirealism) oppose idealism and phenomenalism. They entail the position (or at least are consistent with the

claim) that material objects are real (be they the middle-sized entities of common sense or unobservable entities).

The substantive disagreement between this antirealist tradition and realism is the *sense* of existence. Verificationist antirealism makes the world (or a set of entities) mind-dependent in a more sophisticated sense: What there is in the world is determined by what can be known to exist (verified to exist, rationally accepted as existing, or the like). Hence it forges a logical-conceptual link between what there is in the world and what is affirmed as existing on the basis that it satisfies suitable epistemic conditions. Accordingly, the realist claim of mind-independence should be understood as logical or conceptual independence: What the world is like does not logically or conceptually depend on the epistemic means and conceptualizations used to get to know it. Scientific realism allows for the possibility of a *divergence* between what there is in the world and what is issued as existing by a suitable set of conceptualizations and epistemic conditions. Verificationist antirealism precludes this possibility of divergence a priori by advancing an epistemic conception of truth. No matter what the details of this conception are, the key idea is that truth is conceptually linked with epistemic conditions so tightly that a theory cannot be false even though epistemically justified (because it meets the relevant epistemic condition, for example, being under ideal circumstances theoretically justified or warrantably assertable). Typically, realists honor the possibility of divergence by adopting a non-epistemic conception of truth (the standard candidate for which is the correspondence theory of truth).

Why should scientific realism incorporate the claim of mind-independence? Why, that is, cannot someone who accepts the reality of unobservable entities but regards them as mind-dependent (in the above sense) be a scientific realist? Ultimately at stake in the debate over scientific realism is a robust sense of objectivity, that is, a conception of the world as the arbiter of our changing and evolving conceptualizations of it. Scientific realism honors this conception by claiming that the world is mind-independent. The kernel of its metaphysical thesis is that science is in the business of discovering what a world that is not of our making is like. This thesis implies that if the natural kinds posited by theories exist at all, they exist objectively, that is, independently of our ability to be in a position to know them, verify them, recognize them, etc., and hence that natural kinds, if anything, make scientific theories true. This robust sense of objectivity contradicts verificationist antirealism. It also blocks a number of projectivist or social constructivist views

about science from being realist. In the view of scientific realism, scientific theories and scientific theorizing in general, instead of projecting (or worse, socially constructing) the structure of the world, discover and map out an already structured, mind-independent world.

## SEMANTICS

Let us call the second thesis of scientific realism, the view that scientific theories should be taken at face-value, *semantic realism*. This view too was motivated by problems with verificationism.

Verificationism, at least in its traditional form as defended by the logical positivists, runs together two separate issues: the evidential basis for the truth of an assertion and the semantic relation of reference or denotation. It thereby conflates the issue of what constitutes evidence for the truth of an assertion with the issue of what makes the assertion true. This conflation was the product of concerns about the meaning of theoretical terms. Some empiricists thought that since the meaning of theoretical terms is not given directly in experience, these terms are semantically suspect. Hence, empiricists (even hard-core positivists like Ernst Mach) sought to show that theoretical statements and terms are parasitic on observational statements and terms.

This line of thought led to *reductive empiricism*, which treats theoretical statements as being disguised talk about observables and their actual (and possible) behavior. Interestingly, this view is consistent with the claim that theoretical statements have truth-values, but it understands their truth-conditions reductively: Their truth-conditions can be fully captured in an observational vocabulary. Hence, theoretical statements are ontologically innocuous: They do not refer to unobservable entities, and so imply no commitments to unobservable entities. Despite the heroic efforts of many empiricists (including the early Rudolf Carnap), all attempts to translate theoretical terms into observational terms have patently failed. As a result, empiricism became liberal. It admitted that theoretical terms and statements have excess content that cannot be fully captured by any reference to observable entities and phenomena.

If evidence-conditions and truth-conditions are kept apart, verificationism loses its bite. Semantic realism, simply put, says that there should not be two semantic standards, one for observational statements and another for theoretical ones. Observational statements, as well as theoretical statements, are true if and only if their truth-conditions obtain. Hence, theoretical terms, no less than observational terms, have putative factual reference. If

theoretical statements cannot be given truth-conditions in an ontology that dispenses with theoretical entities, a full and just explication of scientific theories simply requires commitment to irreducible unobservable entities, no less than it requires commitment to observable entities.

*Instrumentalism* claims that theories should be seen as (useful) instruments for organizing, classifying, and predicting observable phenomena. So the “cash value” of scientific theories is fully captured by what theories say about the observable world. Faced with the semantic realist challenge that theoretical assertions are meaningful and purport to describe an unobservable reality, instrumentalism took refuge in Craig’s theorem and claimed that theoretical commitments in science are dispensable: Theoretical terms can be eliminated en bloc without loss in the deductive connections between the observable consequences of the theory. If this is so, then the very question of whether theoretical terms can refer to unobservable entities evaporates. This challenge led Carl Hempel (1958) to formulate what he called “the theoretician’s dilemma.” If the theoretical terms and the theoretical principles of a theory do not serve their purpose of a deductive systematization of the empirical consequences of a theory, then they are dispensable (unnecessary). But by Craig’s theorem, even if they do serve their purpose, they can still be dispensed with. Hence, the theoretical terms and principles of any theory are dispensable.

Is the theoretician’s dilemma compelling? Note first that the very idea of this dilemma rests on a sharp distinction between theoretical terms and observational ones. This dichotomy was severely challenged in the 1960s, when Pierre Duhem’s view that all observation is theory-laden resurfaced. Along with it came the view that, strictly speaking, there are no observational terms. But even if the dichotomy is accepted, instrumentalism based on Craig’s theorem collapses. It is implausible to think of theories as establishing only a deductive systematization of observable phenomena. Theories also offer inductive systematizations in the sense that theories can be used to establish inductive connections among observable phenomena: They function as premises in inductive arguments and, together with other premises concerning observable phenomena, yield conclusions that refer to observable phenomena. Seen as aiming to establish inductive connections among observables, theories are indispensable. There followed a battery of indispensability arguments, fostered by Sellars (1963) and Quine (1960) among others, suggesting that theoretical terms are indispensable in any attempt to formulate a powerful

and efficacious system of laws and to explain why observable entities obey the empirical laws they do.

Semantic realism opposes both instrumentalism and reductive empiricism. It renders scientific realism an “ontologically inflationary” view. Understood realistically, theories admit of a literal interpretation, that is, an interpretation according to which the world is populated by a host of unobservable entities and processes. Semantic realism is not contested any more. All sides of the debate take theoretical discourse to be irreducible and contentful. It should be clear from the above discussion, however, that making semantic realism the object of philosophical consensus was no trivial feat.

## EPISTEMOLOGY

Let us call the third thesis of scientific realism *epistemic optimism*. Its thrust is that science can and does deliver theoretical truth no less than it can and does deliver observational truth. One can grant semantic (even metaphysical) realism and yet remain epistemically skeptical or agnostic toward scientific theories. This agnostic stance has appealed to empiricists who have come to terms with the collapse of instrumentalism and reductive empiricism. An argument for the realist interpretation of scientific theories is not ipso facto an argument for *believing* in the existence of the entities those theories posit and in the truth of what they say of them.

Can the epistemic thesis be avoided? Some realists, notably Alan Musgrave (1999), think that scientific realism is an exclusively axiological thesis: Science aims for true theories. There is clear motivation for this axiological approach: Even if *all* theories scientists ever came up with were false, scientific realism would not thereby be threatened. There are, however, inevitable philosophical worries about the axiological characterization of realism. First, it seems rather vacuous. Realism is rendered immune against the serious criticism stemming from the empirical claim that science has a poor record in tracking the truth. Second, aiming at a goal (truth) whose achievability by the scientific method is left unspecified makes the supposed regulative role of the goal totally mysterious. Finally, we lose all the excitement of the realist claim that science engages in a cognitive activity that pushes back the frontiers of ignorance and error. Other realists, notably Jarrett Leplin (1997), do take the epistemic thesis to be part of scientific realism, but argue for a minimal or thin version of it: There are possible empirical conditions that would warrant attributing some measure of truth to theories. The problem with this minimal account is that, in the end, it cannot provide a rational or warranted basis

for belief in the unobservable entities posited by science (and the assertions made about them).

Naturally, the scope of the epistemic thesis need not (and should not) be universal. Scientific realists need not take current science uncritically. They need not commit themselves to everything that current theories assert. They can have a differentiated attitude toward the theoretical constituents of modern science: Some of them are better supported by the evidence than others; some play an indispensable explanatory role, while others do not; some contribute to the successes of theories, while others do not. But we should not lose sight of the general philosophical issue at stake, which is this: Are there good reasons to believe that science cannot achieve theoretical truth? That is, are there good reasons to believe that, given that we understand the theoretical statements of scientific theories as genuine propositions, we can never be in a warranted position to claim that they are true (or at least, more likely true than false)? The *epistemic thesis* denies that there are such good reasons and defends the claim that the ampliative-abductive methods of science are reliable and can justify/support theoretical assertions. Hence, science has succeeded in tracking truth. To be sure, this success requires a certain amount of epistemic luck: It is not a priori true that science has been, or has to be, successful in truth tracking. If science does succeed in truth tracking, this is a radically contingent fact about how the world is and how science and its method have managed to latch onto it.

The prime argument in favor of the epistemic thesis has come to be known as “the no-miracles argument.” It is an abductive argument, or inference to the best explanation. Jack Smart (1963) argued against instrumentalists that they must believe in cosmic coincidence. On the instrumentalist view of theories, a vast number of ontologically disconnected observable phenomena are “connected” only by a purely instrumental theory: These phenomena just *happen* to be related to one another in the way suggested by the theory. Scientific realism, in contrast, leaves no space for a cosmic-scale coincidence: It is *because* theories are true and *because* the unobservable entities posited by them exist that the phenomena are related to one another as they are. Smart’s key point was that scientific realism (and its concomitant view of science) should be accepted because it offers the best explanation of why the observable phenomena are as scientific theories predict them to be.

Hilary Putnam (1975) and Richard Boyd (1973) argued that inference to the best explanation is the very method scientists use to form and justify their beliefs in

unobservable entities, and that realism should be seen as an overarching empirical hypothesis deriving support from the fact that it offers the best explanation of the success of science. The no-miracles argument found pithy expression in Putnam’s encapsulation: “The positive argument for realism is that it is the only philosophy that does not make the success of science a miracle” (1975, p. 73). A key element of the realists’ epistemic optimism comes from the fact that some theories, because they yield novel predictions, can serve as “prophets for us,” as Duhem put it. Only on a realist understanding do novel predictions about phenomena come as no surprise.

How exactly does the no-miracles argument support the epistemic thesis? Though this issue has been extensively debated, the role of the no-miracles argument in the realism debate is quite complex. To a good approximation, the argument should be seen as a grand inference to the best explanation. It is a philosophical argument that aims to defend the reliability of scientific methodology in producing approximately true theories. The argument proceeds in two steps. The first is that we accept as approximately true the theories that are implicated in the (best) explanation of the *instrumental* reliability of first-order scientific methodology. The second step is that since these theories have typically been arrived at by means of inference to the best explanation, such inference is reliable. The main strength of the no-miracles argument rests on the first part of the argument. Coming after more concrete types of explanatory reasoning that occur all the time in science, the argument suggests that it is reasonable to accept certain theories as approximately true, at least as concerns their components that guided predictions. These successful instances of explanatory reasoning in science provide the basis for the grand abductive argument. However, the no-miracles argument is not just a generalization over the scientists’ abductive inferences. Although itself an instance of the method that scientists employ, it aims at a much broader target, specifically, to defend the thesis that inference to the best explanation (a *type* of inferential method) is reliable. This relates to the second step of the argument. What makes the no-miracles argument distinctive as an argument for realism is that it defends the claim that theoretical truth is achievable. The second step of the argument seeks to secure this claim. It is reasonable to believe that abductive reasoning is reliable, since it tends to generate approximately true theories.

There are two challenges to scientific realism. The first relies on the claim that the evidence underdetermines theories and is discussed in a separate entry. The

second argument is the so-called pessimistic induction. As Larry Laudan (1984) pointed out in developing this argument, the history of science is replete with theories that were once considered empirically successful and fruitful but that turned out to be false and were abandoned. If the history of science is a wasteland of aborted best theoretical explanations of the evidence, then it might well be that current best explanatory theories will travel the route to this wasteland in due course. The best defense of realism against the pessimistic induction has been to try to reconcile the historical record with some form of realism. To do this, realists need to be more selective in what they are realists about.

A claim that emerged with some force in the 1990s is that theory-change is not as radical and discontinuous as the opponents of scientific realism have suggested. Realists such as Philip Kitcher (1993) and Stathis Psillos (1999) have sought to ferret out the theoretical components of abandoned scientific theories that essentially contributed to their successes, separate them from other idle components, and demonstrate that the components making essential contributions to the empirical success of the theories were retained in subsequent theories of the same domain. In such a scenario, the fact that our current best theories may be replaced by others does not necessarily undermine scientific realism. All that such evolution shows is that we cannot get at the truth all at once, and that our judgments from empirical support to approximate truth should be more refined and cautious in that they should commit us only to the theoretical components that enjoy evidential support and contribute to the empirical successes of the theory. Realists ground their epistemic optimism on the fact that newer theories incorporate many theoretical components of their superseded predecessors, especially those components that have led to empirical successes. The substantive continuity in theory-change suggests that a rather stable network of theoretical principles and explanatory hypotheses has emerged, survived revolutionary changes, and become part and parcel of our evolving scientific image of the world.

Faced with the challenge of the pessimistic induction, other realists have sought to weaken realism. There have been two prominent strategies for weakening realism. The first is to opt for structural realism, and the second is to opt for entity realism. Structural realism, defended by John Worrall (1989), capitalizes on the fact that despite the radical changes at the theoretical level, successor theories have tended to retain the *mathematical structure* of their predecessors. It argues that theories can

successfully represent the structure of the world even when they are wrong about the entities they posit. Despite its initial appeal, it turns out that this particular position is very difficult to defend. For one, the distinction between the mathematical structure of the theory and its theoretical content is not as clear-cut as it initially seems. For another, even if a sharp distinction is granted, it turns out that structural realism collapses the difference between the claim that a theory is true and the claim that it is empirically adequate.

Entity realism, defended by Nancy Cartwright (1983) and Ian Hacking (1983), accepts the existence of all sorts of unobservable entities but denies the truth of the theories in which descriptions of these entities are embedded. A major motivation for entity realism comes from laboratory life. Experimenters have good reasons to believe in specific unobservable entities, not because they accept the relevant theories, it is claimed, but rather because they *do* things with these entities. If these entities did not exist, the phenomena of the laboratory would be inexplicable. But can one be a realist about theoretical entities without also being a realist about the theories? In a sense, one can. For posited entities survive theory-change. For instance, scientists accept the existence of electrons even though their theoretical views about what electrons are have changed. So it appears that we can know *that* the electron is, even though we may not know *what* it is. But this cannot be fully right. We cannot assert that electrons are real, that is, that electrons are part and parcel of the furniture of the world, without also asserting that they have *some* of the properties attributed to them by our best scientific theories. So entity realism cannot be fully divorced from theory realism. In any case, the very same inferential process (inference to the best explanation) is involved in accepting the reality of an entity and in accepting the approximate correctness of some theoretical description of it.

## SCIENTIFIC REALISM AND EMPIRICISM

Bas van Fraassen (1980) fostered a rivalry between scientific realism and empiricism with his influential doctrine of constructive empiricism. According to this view about science, (a) science aims at empirically adequate theories, and (b) acceptance of scientific theories involves belief only in their empirical adequacy (though acceptance involves more than just belief; it also involves commitment to the theory). Van Fraassen took realism to be, by and large, an *axiological* thesis: The aim of science is true theories. He supplemented it with a *doxastic* thesis:



Acceptance of theories implies belief in their truth. Seen in this way, realism and constructive empiricism are rivals. But, of course, a lot depends on whether an empiricist ought to be a *constructive* empiricist. There is no logical obstacle impeding an empiricist (who thinks that all knowledge ultimately stems from experience) from fostering methods that warrant belief in the truth of theories in a way that goes beyond belief in their empirical adequacy, and hence from being a scientific realist. Similarly, there is no logical obstacle impeding an empiricist from being stricter than constructive empiricism, for instance, by claiming that (a') the aim of science is unrefuted theories and (b') acceptance of a theory involves the belief only that it is unrefuted.

Constructive empiricism does set the boundaries of experience much farther afield than strict empiricism, and since what empiricism is, is not carved in stone, there is no logical obstacle to setting the boundaries of experience (that is, the reach of legitimate applications of scientific method) even farther afield, as realists demand. Indeed, as Hans Reichenbach (1938) noted, the key question is what kinds of methods are compatible with empiricism. Even if we grant, as we should, that all factual knowledge starts with experience, the boundaries of experience depend on the warrants of the methods employed. It is perfectly compatible with empiricism to accept ampliative methods and to accept the existence of unobservable entities on their basis. So there is no incompatibility between being an empiricist and being a scientific realist.

Van Fraassen tied empiricism to a sharp distinction between observable and unobservable *entities*. This, to be sure, is a step forward from the more traditional empiricist distinction between observational and theoretical terms and predicates. Drawing the distinction in terms of entities allows the description of observable entities to be fully theory-laden. Yet, van Fraassen insisted, even theoretically described, an entity does not cease to be observable if a suitably placed observer *could* perceive it with the naked eye.

Long before van Fraassen, Grover Maxwell (1962) denied this entity-based distinction, arguing that observability is a vague notion and that, in essence, *all* entities are observable under suitable circumstances. He based this view on the claim that "observability" is best understood as detectability by some means. If observability is thus understood, there are continuous degrees of observability, and hence there is no natural and nonarbitrary way to draw a line between observable and unobservable *entities*. Rebutting Maxwell's argument requires that

naked-eye observations (which are required to tell us which entities are *strictly* observable) form a special kind of detection qualitatively set apart from any other way of detecting the presence of an entity (for example, with a microscope). Be that as it may, the issue is not whether the distinction between observable and unobservable entities can be drawn but what its epistemic relevance is: Why should the observable/unobservable distinction define the border between what is epistemically accessible and what is not?

In the end, scientific realism is better than constructive empiricism because (1) it does not rely on a distinction of dubious epistemic significance, specifically, the observable/unobservable distinction, (2) it offers a better explanation of the empirical successes of science, and (3) it tallies better with actual scientific practice.

**See also** Realism; Underdetermination Thesis, Duhem-Quine Thesis.

### Bibliography

- Boyd, Richard. "Realism, Underdetermination, and the Causal Theory of Evidence." *Noûs* 7 (1973): 1–12.
- Cartwright, Nancy. *How the Laws of Physics Lie*. Oxford, U.K.: Clarendon Press, 1983.
- Devitt, Michael. *Realism and Truth*. 2nd, rev. ed. Oxford: Blackwell, 1991. 1st ed., 1984.
- Duhem, Pierre. *The Aim and Structure of Physical Theory*. Translated by Philip P. Wiener. Princeton, NJ: Princeton University Press, 1954. First published in 1906.
- Fine, Arthur. *The Shaky Game*. Chicago: University of Chicago Press, 1986.
- Hacking, Ian. *Representing and Intervening*. Cambridge, U.K.: Cambridge University Press, 1983.
- Hempel, Carl. "The Theoretician's Dilemma: A Study in the Logic of Theory Construction." In *Concepts, Theories, and the Mind-Body Problem*, edited by Herbert Feigl, Michael Scriven, and Grover Maxwell. Minnesota Studies in the Philosophy of Science, no. 2. Minneapolis: University of Minnesota Press, 1958.
- Kitcher, Philip. *The Advancement of Science*. Oxford, U.K.: Oxford University Press, 1993.
- Laudan, Larry. *Science and Values*. Berkeley: University of California Press, 1984.
- Leplin, Jarrett. *A Novel Defence of Scientific Realism*. Oxford, U.K.: Oxford University Press, 1997.
- Maxwell, Grover. "The Ontological Status of Theoretical Entities." In *Scientific Explanation, Space, and Time*, edited by Herbert Feigl and Grover Maxwell. Minnesota Studies in the Philosophy of Science, no. 3. Minneapolis: University of Minnesota Press, 1962.
- Musgrave, Alan. *Essays on Realism and Rationalism*. Amsterdam: Rodopi, 1999.
- Psillos, Stathis. *Scientific Realism: How Science Tracks Truth*. London: Routledge, 1999.

- Putnam, Hilary. *Mathematics, Matter and Method*. Cambridge, U.K.: Cambridge University Press, 1975.
- Quine, Willard Van Orman. *Word and Object*. Cambridge, MA: MIT Press, 1960.
- Reichenbach, Hans. *Experience and Prediction: An Analysis of the Foundations and the Structure of Knowledge*. Chicago: University of Chicago Press, 1938.
- Sellars, Wilfrid. *Science, Perception, and Reality*. Atascadero, CA: Ridgeview, 1963.
- Smart, J. J. C. *Philosophy and Scientific Realism*. London: Routledge and Kegan Paul, 1963.
- Van Fraassen, Bas C. *The Scientific Image*. Oxford, U.K.: Clarendon Press, 1980.
- Worrall, John. "Structural Realism: The Best of Both Worlds?" *Dialectica* 43 (1989): 99–124.

**Stathis Psillos (2005)**

## SCIENTIFIC REVOLUTIONS

Largely as the result of Thomas Kuhn's work, the concept of scientific revolution gains an importance in post-positivist philosophy of science that it lacks in the dominant logical empiricist tradition of the twentieth century. Kuhn's notion of scientific revolution becomes wedded to a historical relativism concerning scientific knowledge that many have sought to refute, or overcome with new accounts of knowledge that go beyond positivism and relativism.

### THE CONCEPTION OF SCIENTIFIC REVOLUTION IN TRADITIONAL PHILOSOPHY OF SCIENCE

To set the context for these debates, it is useful to begin with the ordinary concept of scientific revolution and understand why it lacks fundamental epistemological significance in traditional philosophy of science. In ordinary parlance, a scientific revolution is a large-scale change in the fundamental concepts, theories, or methods that scientists in some area of inquiry employ to understand the course of nature (e.g., the Copernican revolution in astronomy). Such a change is also thought to be revolutionary in so far as it provokes similarly dramatic alterations in the way laypeople see the world. As such, the notion is obviously important to historians of science and popular culture. On the other hand, scientific revolution is not a central topic for the tradition of logical positivism (more broadly, logical empiricism) that generates the key figures, problems, and models of philosophy of science for most of the twentieth century.

In this tradition, the aim of philosophy of science is to provide analyses of the standards most vital to science

as the best exemplar of empirical knowledge: the standards of scientific method, confirmation, prediction, falsification, explanation, truth, progress, observation, law, and theory. The philosopher's analyses are supposed to be timeless, normative, universal, non-historical, and non-empirical. To this end, logical empiricists employed the tools of logic and semantics to illuminate the a priori formal structure of all genuine scientific knowledge (such as explanation and confirmation). Science is identified with its most successful theories, which in turn are represented as finished bodies of propositions linked by logical and inferential relations connecting sense experience to the higher reaches of law and theory.

From this perspective, scientific revolutions alter the content of successful theories, but not the logic of scientific rationality and knowledge. Indeed, the empiricist's logical standards (e.g., Carl Gustav Hempel's deductive-nomological model of explanation, prediction and confirmation) provide the grounds for evaluating the scientific revolutions of Copernicus, Galileo, Johannes Kepler, Sir Isaac Newton, and Albert Einstein. This entire development could be reasonably represented as a logical, cumulative progress. On the philosopher's standards, this progress is one in which, for example, better confirmed theories of wider explanatory scope replace lesser predecessors, whose errors are corrected, and whose sound results are preserved and extended by their successors. The history of the best science(s) illustrates but does not alter the logic of scientific knowledge. So understood, the rationality of science makes it possible for humankind's best theories to converge on the truth concerning lawlike regularities in the world of observed phenomena and, perhaps, the underlying, unobservable entities and mechanisms causally responsible for these regularities.

These achievements of logical empiricism gain one of their last, most lucid and systematic reformulations in Hempel's *The Philosophy of Natural Science*. This work appeared in 1966 four years after Kuhn's *The Structure of Scientific Revolution (SSR)*. Of course many philosophers besides Kuhn challenge one or more of the presuppositions of traditional philosophy of science and reshape the debates in the post-positivist period (e.g. William Van Orman Quine, Wilfred Sellars, Norwood Hanson, Stephen Toulmin, Michael Scriven, Nelson Goodman, Paul Feyerabend, Mary Hesse, etc.). But Kuhn's challenge in *SSR* is probably unique in the avalanche of criticisms, rebuttals, and new approaches to the history and philosophy of science that it has provoked for decades. Much of this response focuses on Kuhn's notion of scientific revo-

lution and the incommensurability, relativism, and irrationalism it is taken to imply.

### KUHN'S CONCEPTION OF SCIENTIFIC REVOLUTION

In effect, Kuhn mobilizes a new conception of the history of science, in which scientific revolution is fundamental and its nature contradicts the formal rationality, normativity, universality, logicism, and progressive cumulativity sought by logical empiricists, and still embraced in new forms by contemporary philosophers (e.g., scientific realists). The philosophical thrust of Kuhn's notion of scientific revolution can be tersely expressed as the following claim. It is in the very nature of (a) science that it undergo not simply changes in the content of its theories, but more fundamentally changes in the very language, problems, goals, and standards that (re)define science, the criteria of scientific knowledge, and membership in the scientific community. This sort of change is what Kuhn's conception of scientific revolution implies, an epistemological change in the requirements of scientific knowledge, explanation, proof, and confirmation. The claim that the essence of science is to generate scientific revolutions, in its own epistemological self-definition, seems like a general philosophical claim. But it is not an a priori claim, for Kuhn. Rather the claim is supposed to be justified by showing that it provides the best explanation of the actual development of science, which opens it up to criticism on this score. In any case, this argument gives history a central role in the evaluation of a philosophy of science.

Kuhn's view of scientific development turns on its division into periods of normal science marked by a normative consensus in the scientific community concerning how to conduct inquiry; and periods of scientific revolution, marked by the breakdown of this consensus. Revolutionary periods typically end when the scientific community is redefined on the basis of a new consensus that creates a different framework for normal science. The normative consensus required by normal science involves the existence of a paradigm that all experts accept as the basis of their research. A scientific revolution implies the dissolution of one paradigm and its eventual replacement by another. A paradigm is a concrete solution (e.g., Lavoisier's account of combustion) to a particular problem (why do some substances gain weight in combustion) that members of a scientific community commonly recognize as an exemplar of how to pursue inquiry in a wider domain of phenomena (chemical reactions); phenomena that may prove to be of the same or similar kind as the paradigm first treated. A

group of inquirers only becomes a scientific community when their research generates a paradigm. As the central object of normative consensus, the paradigm guides practitioners in commonly recognizing what counts as a legitimate problem or phenomenon-to-be-explained in the domain of their science. It tells them what concepts, techniques, mechanisms, measurements, and standards must be present for a legitimate solution to the problem, a bona fide scientific explanation of it. Normal science is the research undertaken to articulate and extend the paradigm by solving a host of puzzles that arise in the attempt to reduce ever-wider phenomena to its terms.

In this process, the shared commitments of the scientific community grow and encompass the formulation of theories, laws, basic equations, standards of proof, mathematical techniques, and experimental procedures. In some contexts, Kuhn refers to this entire body of commitments as the paradigm. Normal science allows a cumulative progress of scientific knowledge, but it is progress within the paradigm, relative to its standards of puzzle solving and explanation. Normal science breaks down when the paradigm confronts anomalies. Anomalies are problems that it ought to be able to resolve, but over time cannot, and that motivate some practitioners to represent the problem, or attempt solutions in ways that abandon basic components of the paradigm and the normative consensus underlying the research tradition defined by it. For Kuhn, one of the best examples of scientific revolution is the abandonment of the premodern chemistry of the phlogiston theory and the theory of elective affinity, due to Lavoisier's oxygen theory of combustion and the new compositional paradigm of Daltonian chemistry.

### KUHN'S CONCEPTS OF INCOMMENSURABILITY

Phlogiston chemistry succeeded in explaining many qualitative phenomena with a paradigm that posits the existence and properties of phlogiston (the presence of phlogiston solves the problem of why the metals have common metallic qualities lacking in their ores). But the phlogiston theory explained the combustion of a substance as a loss of phlogiston that implied weight loss. The phenomenon of weight gain in combustion constituted an anomaly for phlogiston theory because despite serious attempts, no phlogiston chemist succeeded in accounting for it within the constraints of this paradigm. As inquirers abandoned different components of the phlogiston paradigm, in order to accommodate the phenomena of combustion, the road was paved for a revolutionary transformation in the very concepts, language,

questions, techniques, data, values, aims, and standards at the heart of chemistry. In *SSR*, Kuhn stresses the discontinuities marked by scientific revolution and advances his most controversial claim that these discontinuities imply incommensurability between the paradigms or theories separated by scientific revolutions. Incommensurability seems to imply that pre- and postrevolutionary theories cannot be compared because there is no common measure to ground comparison. Such a view is at opposite poles from the project of logical empiricists and their heirs to establish a framework of concepts and standards external to particular theories and their history, and capable of grounding critical evaluation, and judgments of cognitive progress.

But there are different lines of argument in Kuhn concerning the sources and implications of incommensurability. Rival theories are said to be incommensurable because (1) they do not share the same language, or conceptual scheme, and the language of one is not translatable into the language of the other, or a neutral observation language; (2) they do not perceive or recognize the same observational data; (3) they do not address or acknowledge the same problems; (4) they do not embrace the same standards of theory-evaluation or the same interpretations of standards; and (5) they do not live in the same world. While all of these claims are present in Kuhn's argument, which of these sources of incommensurability is most basic, or most defensible? How much room does it leave for continuity and commensurability at the other levels of scientific development? These questions raise the issue of what role reasoning plays in Kuhn's conception of scientific revolution, and how large a role is played by psychological and sociological processes.

Kuhn's very notion of a paradigm and a paradigm-change is sociological in so far as it involves the collective mechanisms through which a scientific community builds up and protects a shared allegiance to its norms and social control over who is and is not a member. He characterizes scientists' embrace of a new paradigm in psychological terms as a gestalt-switch, a leap of faith, and a conversion experience. What role, if any, is left for reason (confirmation, proof, prediction, falsification) in scientists' acceptance of (1) a new conceptual scheme; (2) a new domain of observational data; (3) a different agenda of problems; (4) different standards of theory-evaluation; or (5) a novel world? Which of these is the most basic source of incommensurability? Kuhn's readers and critics focus on different strands of this account of scientific revolution and in response, move philosophy of science in different directions.

### THE FIRST WAVE OF CRITICS: INCOMMENSURABILITY AS TOTAL MEANING CHANGE AND EXTREME RELATIVISM

The first influential line of criticism (Scheffler 1967, 1972; Shapere 1964, 1966, 1971) takes Kuhn's notion of scientific revolution to rest on a radical, holistic conceptual relativism and an implausible view of systematic meaning-variance between paradigms and theories. In essence, on this reading, the first alleged source of incommensurability, paradigms' unique untranslatable language of science, is taken to imply all the others, incommensurabilities of data, problems, standards, and worlds. Each scientific paradigm is imprisoned within its own framework of theoretical concepts whose internal relations determine the unique meaning of each concept and all observation terms employed by the paradigm. On this reading of Kuhn, scientific revolutions change the meaning of all concepts employed by the exponents of a paradigm (e.g., planet in the Copernican revolution) and no translation is possible between the rival languages of science.

With no language in common, it is easy to see why Kuhn would also hold that rival paradigms cannot share common observational data, problems, standards, or worlds. But in that case the advocates of rival paradigms cannot communicate or argue and thus their commitments (beliefs, values, etc.) must be explained by nonrational psychological and sociological processes. Furthermore, retrospective evaluations of theories of the sort grounded in the criteria of traditional philosophy of science (degree of confirmation, explanatory scope, etc.) will be impossible; because there will be no neutral language that permits comparisons of their empirical content. Thus Kuhn's concept of scientific revolution leads to a radical incommensurability and extreme relativism, on which every paradigm, or research tradition, is justified on its own terms, and none is any better than another (better confirmed, etc.).

For the first wave of Kuhn's critics, the resulting position of Kuhn's analysis is incoherent and a "reductio" of its own premises. If rival paradigms cannot be compared or communicate in a common language, in what sense are they rivals? With no common subject matter, there is nothing for them to disagree about. In that case, there would be no difference between a shift of paradigms (or scientific revolution) within a scientific discipline (Cartesian to Newtonian physics) and the movement of inquirers from one area of inquiry into an entirely different one (physicists becoming neuroscientists).

Furthermore, Kuhn's notion of anomalies implies that rival paradigms share some common observational data about which they disagree, and which allow comparisons of their empirical content and success. In that case, they must share some concepts or language, undermining the thesis of radical conceptual incommensurability. Finally the holistic conception of scientific meaning depends on a failure to distinguish sense and reference, among other flaws. Even if the reference of a concept changes ("planet" from Ptolemy to Copernicus; "mass" from Newton to Einstein), there may be sufficient stability of connotation to yield commensurability. On the other hand, when the connotation of observational concepts (temperature of a gas) changes, there is often sufficient stability of reference to allow comparison of paradigms' empirical contents. The development of causal theories of reference reinforced the arguments for continuity of reference (Kitcher 1978, Psillos 1999).

This entire line of criticism located the failure of extreme relativism and radical incommensurability within the terrain of philosophy of language and Kuhn's false starts there. It convinced many philosophers of science that whatever its problems, the tradition of logical empiricism had little reason to worry about Kuhn's notions of scientific revolution and incommensurability.

#### INCOMMENSURABILITY AS SHIFTS-IN-STANDARDS

A second reading of Kuhn shifts the focus to the strain of argument that bases incommensurability not on language, but rather on shifts in the epistemic standards or values that accompany scientific revolutions (Doppelt 1978, 1980; Zammito 2004). Such changes transform the criteria of theoretical knowledge and successful inquiry, for the field and scientific community in question. An allegiance to the new standards implicit in a paradigm shift typically involves a redefinition of the domain of problems and observational phenomena most important for any adequate theory to explain. These shifts sometimes generate losses in the problem-solving capacity and explanatory power of science, though the epistemic importance of these losses is evaluated differently on the disparate standards implicit in rival paradigms.

The premodern chemistry of the phlogiston theory and the theory of elective affinity generated solutions to a large number of problems that are eliminated from the domain of phenomena-to-be-explained by the modern chemistry instigated by Antoine-Laurent Lavoisier and John Dalton. It could account for the observable properties of a number of substances, solving the problem of

why metals exhibited common metallic qualities, lacking in their ores, and why metals take on acidic qualities as a result of chemical reactions. While such questions could still be formulated in the nineteenth century, the failure to answer them, to explain the observed qualities of compounds, is not taken as a cognitive defect in Daltonian chemistry; even though empirical success with these phenomena was a central criterion of theoretical knowledge for premodern chemistry. Of course the modern chemistry of Lavoisier and Dalton succeeded in solving a whole range of problems (concerning weight relations and proportions in chemical reactions) that were largely unknown until their work. Still, given Kuhn's "loss-of-data" and "shift-in-standards" arguments concerning scientific revolution, on his view, the Daltonian paradigm is not well characterized as simply offering a better, truer, or more rational account of chemical phenomena than its predecessor. For, the premodern and modern paradigms provided explanations of different sorts of observed phenomena, in accordance with different problem-sets, and in line with different standards of adequacy for chemical theory.

Reading Kuhn's argument in this way generates a more moderate notion of scientific revolution, incommensurability, and relativism than the initial critics identified. The argument is compatible with considerable continuity and overlap across paradigms concerning language, observational data, problems, and even standards. The existence and role of anomalies exhibits such overlap. More generally, this reading is compatible with Kuhn's clear recognition that new paradigms often try to, and succeed at, treating many of the phenomena at the heart of their predecessors, and satisfying some of their standards, as well as their own. What, then, is left of incommensurability and relativism, in moderate form? Is there a moderate form of these doctrines?

On the moderate version of Kuhn, advocates of rival paradigms present good reasons and arguments to one another. But because their disagreement is about the standards of their science, and the strength of reasons is relative to such standards, paradigm debates and shifts (scientific revolution) are often marked by an absence of compelling reasons. Equally scientific and rational inquirers can weight the balance of good reasons in contradictory ways that favor the standards and achievements implicit in their rival paradigms. This moderate notion of incommensurability of reasons generates a distinctive Kuhnian version of the underdetermination of theory by evidence. Antirealists often argue that the observational implications of a theory do not confirm the

truth of the theory. Because one can always imagine another theory  $T_1$ , incompatible with  $T$ , with the same confirmed observational implications; the two theories are empirically equivalent but cannot both be true. Realists reply that evidence and confirmation involve more than the mere logical consequences derivable from the theory. Confirmation of a theory by evidence for many realists requires that the theory provide the best explanation of the evidence, in virtue of its simplicity, accuracy, explanatory scope, fruitfulness, plausibility, and unification. Kuhn acknowledges the universality of such epistemic values in science. But he argues that shifts-in-paradigms change the criteria governing their application and their relative importance in determining the best explanation. Premodern and modern chemistry both valued unifying explanation, but embraced different standards concerning what sorts of phenomena required unified explanation.

If theory, or paradigm choice, is underdetermined by evidence, and good reasons, due to Kuhn's shift-in-standards claims, reason (scientific method) alone will not explain scientific revolution. Without glorifying irrationalism or mystical conversion, Kuhn can vindicate the relevance of psychological and sociological factors to explain which particular scientific considerations, in an ocean of conflicting reasons, prove compelling to the practitioners who accept a new paradigm, and why. Moderate relativism thus asserts that scientific development involves revolutions in which a new paradigm triumphs, even though it entails some losses in problem-solving capacity, and is no more rational to accept than its predecessor(s), given the different standards at play in the historical context.

### CRITICS OF MODERATE RELATIVISM

This more moderate version of Kuhn's conception of scientific revolution moves its evaluation away from the philosophy of language onto the terrain of epistemological argument. Various critics of Kuhn's shift-of-standards relativism advance arguments based on the existence of external standards, piecemeal bootstrap scientific rationality, naturalist epistemology, and scientific realism (discussed, in turn, below). In the spirit of logical empiricism, some critics argue that Kuhn's emphasis on internal paradigm-specific standards is fully compatible with the existence of external, universal, and non-relative standards of scientific rationality and progress; such as predictive accuracy, explanatory scope, simplicity, completeness, empirical success, unifying power, and the like (Scheffler 1967). Isn't the existence of such independ-

ent standards what makes rational debate between exponents of rival paradigms possible and indeed intelligible as such to us today (Siegel 1980, 1987)?

Kuhn fully embraces the existence of such universal epistemic considerations (empirical success, etc.) in science. But he argues that they function as broad, abstract values of scientific inquiry, whose actual contents are transformed by scientific revolutions. In effect, he takes a moderate relativism of internal standards to imply a relativity of external standards to paradigms. But this is not supposed to be an a priori claim about scientific development. Kuhn's studies of normal science, revolution, and scientific debate are supposed to show that exponents of rival paradigms apply the aforementioned epistemic values in very different ways, yielding concretely different standards of explanation, simplicity, unification, and even accuracy (what counts as an acceptable measure of experimental deviation of prediction from observed result).

But does Kuhn's moderate relativism concerning the role of reasons and standards in scientific revolution imply any relativism concerning long-run scientific progress? The tradition of logical empiricism concerns the context of justification, not discovery. As long as there are external standards of theory-assessment sufficient to establish that science overall attains cognitive progress, Kuhnian short-run losses in problem-solving and standards need not imply any global relativism. As Kuhn himself observes these losses are often recouped in the long run. Though the chemical revolution initiated by Lavoisier abandons the effort to explain the qualities of compounds, these problems are taken up and resolved in twentieth century science. Newtonians first accepted and later abandoned the Aristotelian and Cartesian standards requiring a mechanical explanation of motion, thus gravity (no action-at-a-distance). Einsteinian physics produces an explanation of gravity without any loss to the data and problems handled by Newtonian science.

Kuhn explicitly claims that scientific development exhibits progress in the sense that there are dramatic increases in the number, range, variety, and accuracy of its problem-solutions (even if it is not consistently cumulative, step by step). Another critic seizes on problem-solving effectiveness as the way to accommodate Kuhn's historical insights while overcoming his relativism concerning scientific rationality and equivocations about cognitive progress (Laudan 1977). He seeks to establish an external standard of problem-solving effectiveness with a theory-neutral calculus for identifying, counting, and weighing the various empirical and conceptual prob-

lems tackled, solved, and unsolved in rival or successive research traditions. This account follows Kuhn's historicism in allowing that rival research traditions (a looser, more flexible concept than paradigm) are often committed to different problems, different standards of solution, different criteria for individuating, counting, and weighing important kinds of conceptual and empirical problems.

By accepting the historical relativity of problems, solutions and standards, the externalist model of maximal problem-solving effectiveness runs the risk of collapsing into a Kuhnian moderate relativism concerning the rationality of scientific change and cognitive progress. For example, on the externalist model, the objective importance of a problem (how much it affects a tradition's problem-solving effectiveness) is elevated if rivals tackle and solve it. Against this criterion, the problem solutions taken to be most important in establishing the chemistry first, of Lavoisier, and later, of Dalton, address phenomena that were largely unknown to premodern chemists (e.g., the alchemists) and thus should enjoy less epistemic weight than they were accorded and needed in the making of the chemical revolution. Once external standards are historicized, relativism threatens.

A second critique of Kuhn's notion of scientific revolution follows Kuhn in rejecting self-sufficient external standards and embracing a historicized account of scientific rationality, but one without relativist implications. These critics argue that there are typically good reasons for altering the standards and goals of scientific inquiry, internal to the historical context of shared beliefs in which the change occurs. If the context of shared belief can provide inquirers with a justification for preferring some standards over others, then paradigm change is in principle entirely rational and explainable by the reasons in its favor, without recourse to psychological and sociological dynamics (Siegel 1980, 1987). Some philosophers adopt a multilevel, piecemeal, and gradualist model of scientific change to show precisely how and why the background context of scientific change provides inquirers with good reasons to make these changes (Laudan 1984, Shapere 1984).

The gradualist model directly challenges Kuhn's holistic historiography of normal and revolutionary science. Normal science is supposed to be ordered by a global framework of tightly interwoven concepts, problems, theories, standards, and aims, such that change of any one component implies alterations in all the others. Scientific revolutions are supposed to imply something like a sudden and wholesale break with the entire frame-

work (extreme incommensurability and relativism), or at least its alleged foundational standard(s) (moderate relativism), and the acceptance of a wholly new one.

Gradualist critics argue that if this is what scientific revolutions are supposed to be, then either there are not any, or very few. The process of rebuilding the framework of scientific inquiry is piecemeal and gradual. Change at one level—whether it is theoretical beliefs, empirical observation, methodological standards, or broad cognitive aims—does not dictate change at all other levels; and no one level is foundational for all the rest (Laudan 1984). On the other hand, change on any one of these levels can be justified by elements of continuity and agreement at other levels, even if we accept the Kuhnian view that there are no sacrosanct or permanent aims and standards with which to anchor justification (Shapere 1984).

To take a well-known example, consider the decision of inquirers during the nineteenth century to abandon an exclusive commitment to the Newtonian standard of inductive generalization, which ruled out the epistemic rationality of using observation to support inference to unobservable entities and processes. The strict empiricist inductive standard of proof was widely thought to be responsible for the great Newtonian achievement and its decisive methodological break with the vacuous, speculative hypotheses of Cartesian physics. But in the eighteenth and nineteenth centuries, scientific practitioners became increasingly interested in explaining well-known electrical, chemical, magnetic, gravitational, optical, and other sorts of observed phenomena.

This set of aims took their inquiries beyond the strictures of the Newtonian empiricist standard. The most successful theories (George Lesage, David Hartley, Roger Boscovitch) of these phenomena posited the existence of an unobservable ether(s) in order to account for them. The scientific credibility of these problem-solutions, turned on a new standard of theory-assessment, the method of hypotheses (hypothetico-deductive reasoning). Scientists like Lesage defended this standard as a sound route to genuine knowledge, alongside inductive empiricism. Some members of the scientific community became increasingly committed to the aim of explaining these phenomena, outside the privileged domain of Newtonian physics, and to the aether theories that realized this aim. These shared commitments provided good reason to defend the method of hypothesis and abandon inductivism as the sole standard of genuine knowledge (Laudan 1981, 1984).

The other theorists are neither practitioners of normal Newtonian science nor participants in a revolution-

ary break with it. They do not question the Newtonian achievement and do not reject the standards and aims associated with it. By justifying a wider standard of inference than Newtonians allowed, aether theorists grounded the empirical success of their theories and enhanced the internal consistency of their commitments. Scottish natural philosophers like Thomas Reid stuck to the Newtonian standard and thus argued that the ether theories could not embody genuine scientific knowledge. If one thinks of the parties to these debates as members of the scientific community, then it is much more loosely structured than the notion of a paradigm implies. Its members have different levels of commitment to the disparate components of the framework of scientific inquiry at the time. The framework itself may exhibit tensions or inconsistencies that different inquirers seek to resolve in different ways. The gradualist model of scientific change exploits cases like this to show how the historical context provides inquirers with good reasons for embracing a new standard of scientific knowledge.

Some philosophers press the gradualist model further to argue for a historical conception of progressive scientific rationality on which reasoning over time produces dramatic improvements in the standards, methods, and goals of good reasoning itself. For example, ether theories are ultimately discredited, and the method of hypothesis is supplanted by more demanding criteria of abduction (e.g., William Whewell's consilience of inductions). Nonetheless, the ether theorists' defense of an inference to unobservables, to account for observed phenomena, improved subsequent scientists' understanding of how knowledge can be achieved and what form it might take.

Scientific development can thus be understood as a process of learning how to learn, one in which reasoning generates progressive historical improvements in the very goals, methods and standards of good reasoning itself (Briskman 1977, Brown 1977, Laudan 1984, Nickles 1993, Shapere 1984, Zammito 2004). Such accounts of scientific rationality are characterized as bootstrap rationality, the internalization of reasons, evolutionary epistemology, or nonrelativist historicism, depending on which version is at issue. This dialectical growth in scientific rationality itself accounts for a feature of science that Kuhn himself acknowledges—the extraordinary increase in the power of science, what it can do, by way of problem-solving effectiveness, prediction, explanation, and control. If scientific development implies the enlargement of one's very capacity to reason, this account blunts the epistemologi-

cal force of Kuhn's notion of scientific revolution (shifts-in-standards, moderate relativism).

## THE TURN TO NATURALISM AND REALISM

A closely related development is the emergence of naturalized epistemology. The project of naturalistic epistemologists is to characterize scientific knowledge and its methods on the basis of empirical inquiry, not historical narrative of any sort. Scientific method can be characterized as whatever processes of inference are in fact most effective and reliable means to the ultimate aims of science. Some normative naturalists treat the history of science as a body of empirical evidence that can be used to determine which scientific aims are in fact realizable, and which methods are most effective in realizing them (Laudan). Reliabilist naturalists appeal to our best current sciences in order to determine which methods or mechanisms of belief-formation are most reliable in producing true beliefs (Goldman 1988). From the naturalists' standpoint, scientific change and new standards are not evaluated by the internal reasons provided by the historical context to the inquirers reasoning in that context. Rather naturalists appeal to external empirical knowledge in order to determine whether reliable and effective methods have been followed and this determination does not depend on the reasons or standards that inquirers themselves employ. From this standpoint, rational change and progress in science are evidenced by increases in the reliability of its methods and theories in generating true beliefs.

This naturalistic turn provides another way of circumventing Kuhn's notion of scientific revolution and the historical relativism (of reasons) it implies. One problem for naturalist epistemology arises from the plurality of aims or values in scientific inquiry, a central point in Kuhn's work. The naturalist cannot be expected to identify effective and reliable methods, or processes, of scientific inquiry, if its aim is left indeterminate. Is the aim explanation or prediction, maximal accuracy or unification, simplicity or completeness, etc? Even if one settles on a unitary aim such as truths about the world (as reliabilists hold), this does not settle the methodological debate between realists and empiricists, or instrumentalists.

If the aim is theoretical truths concerning the unobservable causes of observational regularities, as scientific realists argue, then they may also be correct in treating inference-to-the-best explanation as the most effective and reliable method. If the only realizable aim is exclu-



sively truths at the observational level itself, or instrumental reliability (as empiricists stress), then other methods may be more effective. Indeed, the debate between empiricists and realists is precisely over the reliability of inference-to-the-best explanation as a method of confirming the truth of theories. While there are good arguments on both sides, they are not mainly the sorts of purely empirical considerations that naturalist epistemology speaks to. They are closer to the normative and conceptual disagreements brought to light by Kuhn's conception of scientific revolution (Dopplet 1986, 1990, 2001).

Indeed Kuhn's conception places him squarely on the side of instrumentalists. His conception allows that science exhibits cognitive progress in the sense that our best current theories possess vastly more empirical success, instrumental reliability, and problem-solving effectiveness than their predecessors. For scientific realists, the great empirical success of our best current theories provides compelling evidence that they are true. If they weren't true, so realists argue, their great success would be a miracle (Boyd 1973, 1984, 1992; Putnam 1975, 1978; Psillos 1999). The realist view of theories provides the best explanation of their success. On the other hand, Kuhn takes his conception of scientific revolution to support an uncompromising antirealism. He sometimes claims that a scientific revolution alters the world, or more weakly, the aspects of the world central to scientific perception and inquiry (Hoynigen-Heune 1993). In addition, scientists' standards of success and truth shift in scientific revolution. For these reasons, scientific revolution is supposed to preclude the cognitive progress of theories toward the truth concerning the underlying, unobservable structure of reality.

Between Kuhn's virulent antirealism, and the argument of current scientific realists, there is a fundamentally different view of which features of science are most important to account for. Kuhn's notion of scientific revolution focuses on shifts in standards and aims. Scientific realists emphasize the remarkable success of our best science in realizing the ambitious standards and aims it has. If what is most important to explain is not how science arrived at its current standards and aims, but rather why the best current theories are so successful in realizing them, then scientific realists' account offers a powerful antidote to Kuhn's relativism.

Yet, scientific realists have not been entirely immune to Kuhn's historicism. One of the most influential criticisms of scientific realism stems from a careful consideration of past science (Laudan 1984). The realist appeals to

the truth, or approximate truth of our best theories, to explain their empirical success. But how will the realist explain the fact that many outdated theories (e.g. the luminiferous ether theory of the propagation of light) were also empirically successful but false, to the best of our knowledge. Indeed, doesn't this record of false but successful theories constitute good inductive evidence that our currently most successful theories are also probably false? In response, scientific realists have turned to these historical cases and provided realist accounts of their successes and failures (Psillos 1999). Taking stock of the history, realists seek to narrow the range of truly successful theories, limit the components of theories confirmed by their success, and secure a greater continuity of reference than Kuhnian revolutions allow.

However its merits are finally judged, Kuhn's conception of scientific revolution drove a very fruitful wedge between traditional philosophy of science and historicism. It realigned the relation of philosophy of science both to the history of science, and studies of specific scientific practices, theories, and controversies. This realignment helped bring a fuller range of sciences such as biology into the purview of philosophy of science, where physics once reigned supreme. The debates inspired by Kuhn's work helped generate the new approaches to scientific method, rationality, and progress previously described. All told, there is more than a little irony in the fact that some of the most vocal and relentless critics of Kuhn's notion of scientific revolution ended up learning, and teaching, the most from it. What first appeared to many as Kuhn's revolution of irrationality, later proves to be a central component in a larger process of rethinking the aims and methods of philosophy of science itself.

*See also* Kuhn, Thomas; Scientific Method.

### **Bibliography**

- Barnes, Barry. *T. S. Kuhn and Social Science*. London: Macmillan, 1982.
- Bird, Alexander. *Thomas Kuhn*. Princeton, NJ: Princeton University Press, 2000.
- Boyd, Richard. "The Current Status of Scientific Realism." In *Scientific Realism*, edited by J. Leplin, 41–82. Berkeley: University of California Press, 1984.
- Boyd, Richard. "Constructivism, Realism and Philosophical Method." In *Inference, Explanation and other Frustrations: Essays in the Philosophy of Science*, edited by J. Leplin, 131–198. Berkeley: University of California Press, 1992.
- Boyd, Richard. "Realism, Underdetermination, and a Causal Theory of Evidence." *Nous* 7 (1973).
- Briskman, Leonard. "Historical Relativism and Bootstrap Rationality." *Monist* 60 (1977): 453–472.

- Brown, Harold. "For a Modest Historicism." *Monist* 60 (1977): 540–555.
- Brown, Harold. *Perception, Theory and Commitment. The New Philosophy of Science*. Chicago: University of Chicago Press, 1977.
- Devitt, Michael. "Against Incommensurability." *Australasian Journal of Philosophy* 57 (1979): 29–50.
- Devitt, Michael. *Realism and Truth*. Oxford: Basil Blackwell, 1984.
- Doppelt, Gerald. "Kuhn's Epistemological Relativism: An Interpretation and Defense." In *Relativism: Cognitive and Moral*, 113–146. Notre Dame: University of Notre Dame Press, 1978.
- Doppelt, Gerald. "A Reply to Siegel on Kuhnian Relativism." *Inquiry* 23 (1980): 117–123.
- Doppelt, Gerald. "Laudan's Pragmatic Alternative to Positivism and Historicism." *Inquiry* 24 (1981): 253–271.
- Doppelt, Gerald. "Relativism and Recent Pragmatic Conception of Scientific Rationality." In *Scientific Explanation and Understanding: Essays on Reasoning and Rationality in Science*, edited by N. Rescher, 106–142. London: Center for Philosophy of Science Publications in Philosophy of Science, University of Pittsburg and University Press of America, 1983.
- Doppelt, Gerald. "Relativism and the Reticulation Model of Scientific Rationality." *Syntheses* 69 (1986): 225–252.
- Doppelt, Gerald. "The Philosophical Requirements for an Adequate Conception of Scientific Rationality." *Philosophy of Science* 55 (1986): 104–133.
- Doppelt, Gerald. "The Naturalist Conception of Methodological Standards." *Philosophy of Science* 57 (1990): 1–19.
- Doppelt, Gerald. "Incommensurability and the Normative Foundations of Scientific Knowledge." In *Incommensurability and Related Matters*, edited by Paul Hoyningen-Huene and Howard Sankey. Dordrecht: Kluwer Academic, 2001.
- Friedman, Michael. "Remarks on the History of Science and the History of Philosophy." *World Changes: Thomas Kuhn and the Nature of Science*, edited by Paul Horwich. Cambridge, MA: MIT Press, 1993.
- Goldman, Alvin. *Epistemology and Cognition*. Cambridge, MA: Harvard University Press, 1988.
- Giere, Ronald. "History and Philosophy of Science: Intimate Relationship or Marriage of Convenience?" *British Journal for the Philosophy of Science* 24 (1973): 282–297.
- Hempel, Carl. *Philosophy of Natural Science*. Englewood Cliffs, NJ: Prentice-Hall, 1966.
- Horwich, Paul, ed. *World Changes: Thomas Kuhn and the Nature of Science*. Cambridge, MA: MIT Press, 1993.
- Hoyningen-Huene, Paul. *Reconstructing Scientific Revolutions: Thomas S. Kuhn's Philosophy of Science*. Translated by Alexander T. Levine. Chicago: University of Chicago Press, 1993.
- Kitcher, Phillip. "Theories, Theorists and Theoretical Change." *The Philosophical Review* 87 (1978): 519–547.
- Kordig, Carl. *The Justification of Scientific Change*. Dordrecht: Reidel, 1971.
- Kuhn, Thomas. *The Copernican Revolution: Planetary Astronomy in the Development of Western Thought*. Cambridge, MA: Harvard University Press, 1957.
- Kuhn, Thomas. *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press, 1977.
- Kuhn, Thomas. "Logic of Discovery or Psychology of Research?" In *Criticism and the Growth of Knowledge*, edited by I. Lakatos and A. Musgrave, 1–25. Cambridge, U.K.: Cambridge University Press, 1970.
- Kuhn, Thomas. "The Relation between History and the History of Science." In *The Essential Tension*, 127–161. Chicago: University of Chicago Press, 1971.
- Kuhn, Thomas. *The Structure of Scientific Revolution*. 2nd ed. Chicago: University of Chicago Press, 1970.
- Kuhn, Thomas, J. L. Heilbrom, P. Forman, and L. Allen. *Sources for the History of Quantum Physics: An Inventory and Report*. Philadelphia, 1967.
- Lakatos, Imre. "Falsification and the Methodology of Scientific Research Programmes." In *Criticism and the Growth of Knowledge*, 91–197. Cambridge, U.K.: Cambridge University Press, 1970.
- Lakatos, Imre, and Alan Musgrave, eds. *Criticism and the Growth of Knowledge*. Cambridge, U.K.: Cambridge University Press, 1970.
- Laudan, Larry. *Beyond Positivism and Relativism*. Boulder, CO: Westview, 1996.
- Laudan, Larry. "The Demise of the Demarcation Problem." In *Physics, Philosophy and Psychoanalysis*, edited by Robert S. Cohen and Larry Laudan, 111–127. Dordrecht: Reidel, 1983.
- Laudan, Larry. *Progress and Its Problems*. Berkeley: University of California Press, 1977.
- Laudan, Larry. "Progress and Rationality? The Prospects for Normative Naturalism." *American Philosophical Quarterly* 24 (1987): 19–31.
- Laudan, Larry. *Science and Hypothesis*. Dordrecht: Reidel, 1981.
- Laudan, Larry. *Science and Values*. Berkeley: University of California Press, 1984.
- Laudan, Larry. "Two Dogma of Methodology." *Philosophy of Science* 43 (4) (1976): 585–597.
- McMullin, Ernan. "Philosophy of Science and Its Rational Reconstructions." In *Progress and the Rationality of Science*, edited by Gerald Radnitzky, Gunnar Andersson, Robert S. Cohen, and Marx W. Wartofsky, 221–252. Dordrecht: Reidel, 1978.
- Meiland, John. "Kuhn, Scheffler, and Objectivity in Science." *Philosophy of Science* 41 (1974): 179–187.
- Musgrave, Alan. "Logical versus Historical Theories of Confirmation." *British Journal for the Philosophy of Science* 25 (1974): 1–23.
- Newton-Smith, William. *The Rationality of Science*. London: Routledge, 1981.
- Nickles, Thomas. "Scientific Discovery and the Future of Philosophy of Science." In *Scientific Discovery, Logic and Rationality*, 1–59. Dordrecht: Reidel, 1987.
- Nickles, Thomas. *Taking the Naturalist Turn, or How Real Philosophy of Science Is Done*. Chicago: University of Chicago Press, 1993.
- Psillos, Stathis. *Scientific Realism: How Science Tackles Truth*. London and New York: Routledge, 1999.
- Putnam, Hilary. *Mathematics, Matter and Method*. Vol. 1. Cambridge, U.K.: Cambridge University Press, 1975.

- Putnam, Hilary. *Meaning and the Moral Sciences*. London: Routledge, 1978.
- Radnitzky, Gerald, and W. W. Bartley, ed. *Evolutionary Epistemology, Rationality, and the Sociology of Knowledge*. La Salle, IL: Open Court, 1987.
- Rosenberg, Alex. "Normative Naturalism and the Role of Philosophy." *Philosophy of Sciences* 57 (1990): 34–43.
- Sankey, Howard. *The Incommensurability Thesis*. Brookfield, VT: Ashgate, 1994.
- Sankey, Howard. "Methodological Pluralism, Normative Naturalism and the Realist Aim of Science." In *After Popper, Kuhn and Feyerabend: Recent Issues in Theories of Scientific Method*, edited by R. Nola and Howard Sankey, 211–229. Dordrecht: Kluwer, 2000.
- Sankey, Howard. *Rationality, Relativism and Incommensurability*. Brookfield, VT: Ashgate, 1997.
- Scheffler, Israel. *Science and Subjectivity*. Indianapolis, IN: Bobbs-Merrill, 1967.
- Scheffler, Israel. "Vision and Revolution: a Postscript oh Kuhn." *Philosophy of Science* 39 (1972): 366–374.
- Shapere, Dudley. "The Structure of Scientific Revolution." *The Philosophical Review* 73 (1964): 383–394.
- Shapere, Dudley. "Meaning and Scientific Change." In *Mind and Cosmos: Essays in the Contemporary Science and Philosophy*, edited by R. Colodny, 41–85. Pittsburgh: University of Pittsburgh Press, 1966.
- Shapere, Dudley. "The Concept of Observation in Science and Philosophy." *Philosophy of Science* 49 (1982): 485–525.
- Shapere, Dudley. "The Paradigm Concept." *Science* 172 (1971): 706–709.
- Shapere, Dudley. *Boston Studies in Philosophy of Science*. Vol. 78: *Reason and the Search for Knowledge*. Dordrecht: Reidel, 1984.
- Siegel, Harvey. "Epistemological Relativism In Its Latest Form." *Inquiry* 23 (1980): 107–117.
- Siegel, Harvey. *Relativism Refuted*. Dordrecht: Reidel, 1987.
- Thagard, Paul. *Conceptual Revolutions*. Princeton, NJ: Princeton University Press, 1992.
- Zammito, John. *A Nice Derangement of Epistemes*. Chicago: University of Chicago Press, 2004.

Gerald Doppelt (2005)

## SCOT, MICHAEL

(fl. 1217–c. 1240)

Michael Scot was an astrologer, alchemist, and translator of Arabic and Hebrew works into Latin. Born in Scotland late in the twelfth century, he spent most of his active life in Toledo, Palermo, and mainland Italy—perhaps at Rome. He first appears with any degree of certainty at Toledo in 1217, when he finished a translation of al-Bītrōgī's (Alpetragius's) *Liber Astronomiae* (On the spheres). The next certain date is 1220, when he is reported to have completed a Latin translation of Aristotle's *Historia Animalium*, probably at Toledo. He seems to

have become favorably known at the papal court, for he was offered the archbishopric of Cashel in Ireland in 1225. He refused the office because of his ignorance of Gaelic. Probably during this period he produced the translation of Aristotle's *De Caelo et Mundo*, along with several other physical works of Aristotle with their Arabic commentaries by Averroes. It was these commentaries that were to be so influential among the Schoolmen for the next several generations. About 1228, as nearly as can be judged, Scot entered the service of Emperor Frederick II in Sicily, or at his court at Palermo, as his official astrologer. While there, he wrote his compendious *Liber Introductorius*, a general survey of the whole science of astrology, and the *Liber Particularis*, similar in content but much briefer, intended for popular use. He also composed a *Physiognomia*, a general handbook of physiological science. All three works were dedicated to the emperor and brought Scot a wide reputation. From this second, Sicilian period of his life comes the *Abbreviatio Avicenne de Animalibus*, probably done in 1231, in answer to Frederick's request for more scientific information about the animal kingdom. It was also during this period that Scot wrote *De Arte Alchemie* in which he reported having witnessed and himself verified alchemical experiments performed by Arabs, Jews, Spaniards, and north Africans.

Because of his renown many other works have been ascribed to him, such as a commentary on John of Holywood (Sacrobosco) titled *De Sphera* and a Latin translation of Maimonides' *Guide of the Perplexed*, but these attributions lack any proof or, indeed, likelihood. Scot's great contribution remains his work of translation from Arabic and Hebrew sources of Aristotle's zoological works, the work of al-Bītrōgī, the commentaries of Averroes on Aristotle, and the zoological work of Avicenna. Dante Alighieri consigns him to hell as a magician.

**See also** Aristotle; Averroes; Avicenna; Dante Alighieri; Maimonides.

### Bibliography

The biography by J. Wood Brown, *Life and Legend of Michael Scot* (Edinburgh, 1897), must be used with care. For more recent and reliable treatments, see Lynn Thorndike, *History of Magic and Experimental Science during the First Thirteen Centuries of Our Era*, 2 vols. (New York: Macmillan, 1923–1958), Vol. II, pp. 307–337; C. H. Haskins, *Studies in the History of Mediaeval Science* (Cambridge, MA: Harvard University Press, 1927), pp. 272–298; and G. Sarton, *Introduction to the History of Science* (Baltimore: Carnegie Institution of Washington, and Williams and Wilkins, 1931), Vol. II, Ch. 2, pp. 579–582, with a bibliography of editions.

S. Harrison Thomson (1967)

## SCOTISM

Scotism refers variously to (1) a loosely identified body of views thought to be original with or characteristic of John Duns Scotus, (2) a tradition of texts, doctrines, and approaches that traces back to him, and (3) a *via* (“way” or perhaps “school”) that had an institutional presence in the universities of the fifteenth through the seventeenth centuries. In the first sense scholars today speak of “logical Scotism” in the work of authors who perhaps have never heard of Scotus. In both the first and the second senses scholars ask whether Charles Sanders Peirce, who had read Scotus with care but was a fiercely independent thinker, was a Scotist. In the third sense scholars inquire about the presence of Scotism in the arts and theology faculties of particular universities in the sixteenth century and investigate its fortuna. Confusion can arise (and has arisen) from running these together and one should take special care to distinguish them all from the influence of Scotus—which was so pervasive in the later Middle Ages that almost every metaphysician and theologian of note felt obliged to locate himself with respect to the Subtle Doctor.

Scotus died young (perhaps as young as forty-two) and left an enormous number of fertile ideas in various stages of development. His immediate students and followers, particularly those at Paris, among whom one might add William of Alnwick (c. 1275–1333), Antonius Andreas, Nicholas Bonet (?–1360), Francis of Marchia (1290–1344), Francis Meyronnes (c. 1285–after 1328), and Petrus Thomae (c. 1280–c. 1337), took up those they found congenial and developed them in somewhat different directions. Within twenty years of Scotus’s death there had also grown up a number of different explicitly critical responses to his teaching exemplified in the work of Petrus Aureoli (1280–1322) on the one hand and William of Ockham on the other. At least four elements of Scotus’s thought became identified with him in particular: In metaphysics the view that there was an isomorphism between the structure of concepts and the structure of things and the associated postulation of *formalitates*; on the borderline of metaphysics and theology a distinctive argument for the existence and infinity of God; in theology the doctrine that Mary had been conceived immaculately, that is, without the stain of original sin; and also the view that the divine will was the ultimate cause of the truth of all contingent truths. It is not at all clear that any of these doctrines was entirely original with Scotus and so one should be cautious in locating someone who does not self-identify as a follower of Scotus as a Scotist in

either the second or the third senses of the word simply because the person maintains some of them.

In the first half of the fourteenth century it seems to have been the metaphysical doctrines just mentioned that received the most attention. The key concept of a *formalitas* and the closely associated notion of an *haecitas* as a formal principle of individuation attracted the attention of most of the metaphysicians of the period. There even grew up a distinctive genre of treatise *De Formalitatibus* that studied these notions. Scotus’s argument for the existence and infinity of God as developed both in his *Ordinatio* and the treatise *De Primo Principio* became celebrated soon after his death Thomas Bradwardine devoted his enormous *De Causa Dei* to correcting, elaborating, and refining it and there was considerable controversy about it throughout the century. Scotus’s distinctive views about the role of the divine will in the truth of contingent truths also attracted considerable attention. Much of this attention was hostile, but it was intense for all that. In the fifteenth century the doctrine of the immaculate conception, which had been rejected by Thomas Aquinas but maintained by many thinkers including Scotus, Ockham, and Pierre d’Ailly, became associated with Scotus more particularly and by the middle of the sixteenth century, as other alternatives to Thomism faded from the theological scene, it became thought characteristic of Scotism.

The earliest references to a Scotist school or at least to a group of thinkers whom one can identify as such, are, as is quite typical in the Middle Ages, by figures who see themselves as opposed to it. In 1331 Adam Wodeham, no friend of the view, identifies the isomorphism between things and concepts as characteristic of an unnamed group of thinkers who hold it to be the fundamental principle of metaphysics. By 1400 Jean de Gerson (1363–1429) identified a group holding this view as the *formalizantes* and set himself vigorously against it. In the fifteenth century one finds thinkers like John Foxoles both self-identifying as Scotists and attempting to work out histories of the movement with which they identified. Peter of Candia (c. 1340–1410) is a particularly interesting thinker of this period much influenced by Scotus whose work has received modern study.

Scotus’s works were intensively studied throughout the Franciscan order during the fifteenth through the seventeenth centuries and the fortunes of that order considerably influenced his reception. Scotism as a *via* (school) reached its zenith in the seventeenth century. The Irish Franciscans claimed Scotus as their own (in the middle of the seventeenth century the prominent philosopher-the-

ologian John Ponce [1603–1670] even wrote “Scotus Hiberniae restitutus” to prove the point) and under the leadership of Luke Wadding (1588–1657) a team at the Irish college of St. Isidore in Rome prepared an edition of Scotus’s works (Lyons 1639) that has been foundational for all subsequent editions. The considerable intellectual resources of the Franciscan order in the seventeenth century led to interesting philosophical development and debate of which the most celebrated instance is that between Ponce and Bartholomew Mastrius (1602–1673) over the nature of possibility.

**See also** Alexander of Hales; Augustinianism; Bonaventure, St.; Duns Scotus, John; Medieval Philosophy; Peter Lombard; Thomas Aquinas, St.; Thomism.

### Bibliography

- Catto, Jeremy I. “Theology after Wycliffism.” In *The History of the University of Oxford*. Vol. 2, *Late Medieval Oxford*, edited by Jeremy I. Catto and Ralph Evans, 263–280. New York: Oxford University Press, 1992.
- Coombs, Jeffrey S. “The Possibility of Created Entities in Seventeenth-Century Scotism.” *Philosophical Quarterly* 41 (1993): 447–459.
- Courtenay, William *Schools and Scholars in Fourteenth Century England*. Princeton, NJ: Princeton University Press, 1987.
- Cross, Richard. “Medieval Theories of Haeceity.” In *The Stanford Online Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford, CA: Metaphysics Research Lab, Stanford University, 2005.
- Dumont, Stephen “The Scotist of Vat lat 869.” *Archivum franciscanum historicum* 81 (1988): 254–283.
- Gabriel, Astrik, L. “‘Via Antiqua’ and ‘Via moderna’ and the Migration of Paris Students and Masters to the German Universities in the Fifteenth Century.” *Miscellanea Mediaevalia* 9 (1974): 439–483.
- Gelber, Hester. *Logic and the Trinity*. Madison: University of Wisconsin Press, 1974.
- Hoenen, Maarten. “Scotus and the Scotist School: The Tradition of Scotist Thought in the Medieval and Early Modern Period.” In *John Duns Scotus (1265/6–1308)*, by E. P. Bos, 197–210. Amsterdam, Netherlands: Rodopi, 1998.
- Kaluza, Zenon. *Les Querelles doctrinales a Paris*. Bergamo, Italy: Lubrina, 1988.
- Martin, Christopher J. “Theories of Inference and Entailment in the Middle Ages.” PhD diss., Princeton University, 1999.

**Calvin G. Normore (2005)**

## SEARLE, JOHN

(1932–)

John R. Searle was born in Denver Colorado in 1932. He attended the University of Wisconsin (1949–1952), then

Oxford (1952–1959) as a Rhodes Scholar. He earned his PhD (Oxford) in 1959 and went to the University of California Berkeley, where he remained, and where he is Mills Professor of the Philosophy of Mind and Language. Over the past forty years, Searle has been working on a selection of problems in philosophy at three levels of description: mind (the basic level), language (the middle level), and society (the highest level). In each case Searle can be seen as following a certain pattern: he proposes analyses of facts at one level of description in which they cause, are realized in, or constitute, facts at another higher level. Brute facts can count as institutional facts, and some objective brute facts can cause and realize other, subjective, brute facts. Like phenomenological analyses, Searle’s approach is not classically reductive, but there is an explanatory asymmetry: higher level phenomena often are to be explained in terms of lower level phenomena (explaining is not explaining away). However, as contrasted with phenomenology, this procedure does not require that conditions revealed by analysis be revealed in experience.

### MIND, COGNITIVE SCIENCE AND RATIONALITY

Searle (1981) presents the “Chinese Room” argument against “strong artificial intelligence,” the view that mental states are and can be explained by programs running on the brain, by claiming that programs will give you at best the syntax or structure of thoughts, but not their semantics, their intentionality (aboutness). Searle (1985) schematized such intentional states  $S(r)$ , where  $S$  is a psychological mode, such as believing, and  $r$  is a representational or propositional content: that snow is white.  $S$  typically determines the “direction of fit” of the intentional state: beliefs have a mind-to-world direction of fit, intentions and desires have a world-to-mind direction of fit. Together,  $S$  and  $r$  fix conditions of satisfaction. For beliefs this is a truth-condition, for intentions and desires it is a fulfillment-condition. Some intentional states, such as perception, memory, intention, have the added feature of causal self-reference, in that their conditions of satisfaction make reference to their own causal role. All intentional states are linked in a causal and logical network, and function against a background of nonintentional capacities and abilities.

Consciousness, Searle (1992) argued, is not only a unified qualitative experiential state, it is a natural biological phenomena caused by and realized in the brain. Furthermore, according to the “connection principle,” all mental states are either conscious, or available in principle

to consciousness. This principle if correct would rule out many of the kinds of preconscious mental states favored by cognitive science, including linguists' "cognized" principles of language, vision theorists's algorithms for the computation of stereopsis, and philosophers's "functionalist" analysis of intentional states. Furthermore, "cognitivism," the view that brains are computers (digital or connectionist) is mistaken, because being a computer is an observer-relative fact and not an intrinsic feature of the neuroscience of brains. Mental states are ontologically subjective in that they depend on a mind to exist, but they are epistemically objective in that claims about them are true or false independently of opinion.

Searle (2001) claims that human agents can act rationally because they have free choice. There are three potential "gaps" or decision points in the chain leading to free, voluntary action: a gap between having reasons and forming a prior intention to act; a gap between the prior intention and the intention-in-action that causes the movement that counts as the action; and the gap between segments of temporally extended activities—continuing to act. Acting freely involves selecting a reason to act on, and that reason cannot be causally sufficient for the action.

#### LANGUAGE, SPEECH ACTS, AND SOCIETY

According to Searle (1958), Frege was almost right: the use of proper names is backed by descriptive content, not by any particular one, but by a cluster. No particular predication on a name is necessary, but the disjunction of contents is. This doctrine is the target of Saul Kripke's attack on description theories of names. Searle (1969, 1979, 2001) elaborates and defends the idea that speaking a language is a form of rule-governed behavior, and that the semantics of a natural language is to be given in terms of "constitutive" rules for performing speech acts. These rules "regulate" antecedently existing forms of behavior, or "count as" the creation of a new form of behavior, or both. Illocutionary acts, such as asserting that snow is white, typically have the structure  $F(P)$ , where  $F$  is the illocutionary force (assertion) and  $P$  is the propositional content (that snow is white). Sentences typically encode this distinction in devices for indicating the force,  $F$ , of the utterance, and devices for indicating the propositional content,  $P$ , and these devices are governed by constitutive rules for performing the relevant illocutionary and propositional acts. Each illocutionary act has a distinct (illocutionary) point or purpose, which can be used to taxonomize such acts. Searle and Daniel Vanderveken (1985) propose an illocutionary logic in which relations

between illocutionary acts and forces are captured formally.

Many illocutionary acts can be performed explicitly with the performative formula ("I hereby adjourn the meeting"), in which case the speaker makes a self-referential, self-guaranteeing declaration. Illocutionary acts can also be performed indirectly, and nonliterally (metaphor), and the theory of these forms of communication need not appeal to any special principles beyond the constitutive rules for speech acts, general rationality, and Gricean principles of conversation. Viewed from the lower level of intentional states, the performance of a speech act is the mental imposition of conditions of satisfaction on an utterance, which itself satisfies the intention in action to produce that utterance. Hence Searle's recurrent slogan that all meaning involves "imposing conditions of satisfaction on conditions of satisfaction" (Searle 2001, p. 53). Searle (1995) argues that institutions, and social facts in general, are created when agents collectively impose a new status on things that antecedently do not have it, and go on to attach certain functions to that status. Thus, a piece of paper or metal becomes money when exchange value is assigned to it and accepted. The general form of the creation of such institutional facts is: People collectively accept that  $X$  has the power to do  $A$ . Such status-functions can be nested within one another creating tangled hierarchies of social facts and organization—money can pay mortgages for property, and that property can then be inherited. Such "collective intentionality" is basic and cannot be reduced to individual or mutual intentionality.

*See also* Chinese Room Argument.

#### Bibliography

##### PRIMARY SOURCES

- "Proper Names." *Mind* 67 (1958): 166–173.
- Speech Acts: An Essay in the Philosophy of Language*. Cambridge, U.K.: Cambridge University Press, 1969.
- Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge, U.K.: Cambridge University Press, 1979.
- "Minds, Brains and Programs." *The Behavioral and Brain Sciences* 3 (1981): 417–424.
- Foundations of Illocutionary Logic*. With Daniel Vanderveken. Cambridge, U.K.: Cambridge University Press, 1985.
- The Rediscovery of the Mind*. Cambridge, MA: MIT Press, 1992.
- The Construction of Social Reality*. Glencoe, IL: Free Press, 1995.
- Rationality in Action*. Cambridge, MA: MIT Press, 2001.
- Consciousness and Language*. Cambridge, U.K.: Cambridge University Press, 2002.

## SECONDARY WORKS

Grewendorf, Guenther, and Georg Meggle, eds. *Speech Acts, Mind and Social Reality*. Dordrecht, Netherlands: Kluwer, 2002.

Lepore, Ernest, and Robert van Gulick, eds. *John Searle and His Critics*. Malden, MA: Blackwell, 1991.

Smith, Barry, ed. *John Searle*. Cambridge, U.K.: Cambridge University Press, 2003.

**Robert M. Harnish (2005)**

## SECONDARY QUALITIES

See *Primary and Secondary Qualities*

## SECOND ORDER LOGIC

Second-order logic is the extension of first-order logic obtained by introducing quantification of predicate and function variables. A first-order formula, say  $Fxy$ , may be converted to a second-order formula by replacing  $F$  with a dyadic relation variable  $X$ , obtaining  $Xxy$ . Existential quantification yields  $\exists X Xxy$ , which may be read “*there is a relation that  $x$  bears to  $y$* .” In general relation variables of all arities are admissible. Similarly, quantifiable function variables may be introduced.

## SEMANTICS FOR THE SECOND-ORDER LOGIC

A structure, with non-empty domain  $D$ , for a second-order language includes relation domains  $\text{Rel}_n(D)$  and function domains  $\text{Func}_n(D)$ . In general  $\text{Rel}_n(D) \subseteq \mathbf{P}(D^n)$ , where  $\mathbf{P}(D^n)$  is the power set of  $D^n$ . Similarly, the function domains  $\text{Func}_n(D)$  are subsets of the collection of  $n$ -place total functions on  $D$ . Such second-order structures are called *Henkin* or *general* structures. If  $X$  is an  $n$ -place relation variable, a formula  $\exists X\varphi(X)$  is *true* in a Henkin structure  $M$  if there is an  $n$ -place relation  $R \in \text{Rel}_n(D)$  such that  $\varphi(X)$  is true in  $M$  when  $X$  has the value  $R$ . There is a similar definition for formulas of the form  $\forall X\varphi(X)$  and for formulas with quantified function variables. A formula  $\varphi$  is a *Henkin semantic consequence* of a set  $\Delta$  of formulas if  $\varphi$  is true in all Henkin models of  $\Delta$ .

The relation domain  $\text{Rel}_n(D)$  need not contain all subsets of  $D^n$ . If  $\text{Rel}_n(D) = \mathbf{P}(D^n)$  for each  $n$ , we say that each relation domain is *full* (similarly for function domains) and that the structure is *full*, *standard* or *principal*. Second-order logic restricted to full structures is called *full* or *standard* second-order logic. A formula  $\varphi$  is a *full semantic consequence* of a set  $\Delta$  if  $\varphi$  is true in all full

models of  $\Delta$ . A formula is *valid* if it is true in all full structures.

In Henkin semantics, the Completeness, Compactness and Löwenheim-Skolem Theorems hold because Henkin structures can be reinterpreted as many-sorted first-order structures. This yields Henkin’s Completeness Theorem: There exists a deductive system DS such that if  $\varphi$  is a Henkin consequence of axioms  $\Delta$  then there is a *deduction* of  $\varphi$  from  $\Delta$  using the rules of DS. For further details, see Shapiro 1991, Shapiro 2001, or van Dalen 1994.

## EXPRESSIVE POWER

Following Gottfried Leibniz, we may define “ $x = y$ ” as “any property of  $x$  is a property of  $y$ .” The corresponding second-order definition  $\forall x\forall y(x = y \leftrightarrow \forall X(Xx \rightarrow Xy))$  is valid. In contrast with first-order logic, there are *categorical* second-order theories with infinite models: All full models are isomorphic. For example, let  $\Delta$  be the theory with axioms  $\forall x(s(x) \neq 0)$ ,  $\forall x\forall y(s(x) = s(y) \rightarrow x = y)$  and  $\forall X[(X0 \wedge \forall x(Xx \rightarrow Xs(x))) \rightarrow \forall xXx]$ . Any full model of  $\Delta$  is isomorphic to the structure  $(\mathbf{N}, 0, S)$ , where  $\mathbf{N}$  is the set of natural numbers and  $S$  the successor operation. So, the Löwenheim-Skolem Theorems fail in full second-order logic. Consider the theory  $\Delta \cup \{c \neq 0, c \neq s0, c \neq ss0, \dots\}$ , with  $c$  a constant. This theory has no full model, but any finite subset of it has a full model. So the Compactness Theorem fails, too.

Extending  $\Delta$  with the recursion axioms for addition and multiplication, we obtain the theory  $\text{PA}_2$  whose unique full model up to isomorphism is the natural number structure  $(\mathbf{N}, 0, S, +, \times)$ . Similarly there is an axiom system whose unique full model up to isomorphism is the ordered field of real numbers,  $(\mathbf{R}, 0, 1, +, \times, <)$ . More generally there exist second-order formulas expressing cardinality claims inexpressible in first-order logic. The most striking example concerns the Continuum Hypothesis (CH), which says that there is no cardinal number between  $\aleph_0$  and  $2^{\aleph_0}$ . Results due to Kurt Gödel and Paul Cohen imply that the Continuum Hypothesis is independent of standard axiomatic set theory (ZFC). But there is a second-order formula  $\text{CH}^*$  which is valid just in case CH is true.

If we augment  $\text{PA}_2$  with inference rules for the second-order quantifiers and the monadic comprehension scheme  $\exists X\forall x(Xx \leftrightarrow \varphi)$ , we obtain axiomatic second-order arithmetic,  $Z_2$ . (See Simpson 1998 for a detailed investigation of  $Z_2$  and its subsystems.) One may construct a Gödel sentence  $G$ , true just in case  $G$  is not a theorem of  $Z_2$ . Now, all full models of  $Z_2$  are isomorphic to

( $\mathbb{N}, 0, S, +, x$ ). So an arithmetic sentence  $\phi$  is true just in case  $\phi$  is a full semantic consequence of  $Z_2$ .  $G$  is thus a full semantic consequence of  $Z_2$  but not a theorem of  $Z_2$ . The Completeness Theorem therefore fails; there is no sound and complete, recursively axiomatized, deductive system for full second-order logic. Indeed the set of second-order validities is not recursively enumerable. For further details see Shapiro 1991, Shapiro 2001, or Enderton 2001.

### IS SECOND-ORDER LOGIC LOGIC?

Second-order comprehension has the form  $\exists X \forall x_1 \dots \forall x_n (Xx_1 \dots x_n \leftrightarrow \phi)$ . Should such existential axioms count as logical? Does this violate the *topic-neutrality* of logic? W. V. Quine argued that second-order logic is “set theory in sheep’s clothing” because “set theory’s staggering existential assumptions are cunningly hidden ... in the tacit shift from schematic predicate letter to quantifiable variable” (Quine 1970, p. 68). Another reason for not counting second-order logic as logic is that the full semantic consequence relation does not allow a complete proof procedure.

In reply George Boolos pointed out that the obvious translation from second-order formulas to first-order set-theoretic formulas does not map valid formulas to set-theoretic theorems. For example  $\exists X \forall y Xy$  is valid, while  $\exists x \forall y (y \in x)$  is refutable in axiomatic set theory. Furthermore  $\exists X \exists x \exists y (Xx \wedge Xy \wedge x \neq y)$  is not valid, and so “second-order logic is not committed to the existence of even a two-membered set” (Boolos 1975 [1998], pp. 40–41). Furthermore first-order logic does have a complete proof procedure, but the set of first-order validities is undecidable (Church’s Theorem), while the monadic fragment is decidable. So why is completeness used to draw the line between logic and mathematics rather than decidability?

### THE INTERPRETATION OF SECOND-ORDER VARIABLES.

George Boolos (1984, 1985) has provided monadic second-order logic with a novel interpretation: the plural interpretation. Certain natural language locutions that receive monadic second-order formalizations are perhaps better analysed as instances of plural quantification. For example the Geach-Kaplan sentence, “Some critics admire only one another,” may be formalized as  $\exists X (\exists x Xx \wedge \forall x \forall y (Xx \wedge Axy \rightarrow x \neq y \wedge Xy))$ . This formula is non-first-orderizable (not equivalent to a first-order formula containing just the predicates  $A$  and  $=$ ). According to the usual interpretation, its truth implies the existence of a

collection. The plural interpretation reads “There are some [critics] such that, for any  $x$  and  $y$ , if  $x$  is one of them and admires  $y$ , then  $y$  is not  $x$  and  $y$  is one of them.” Rather than asserting the existence of a collection, this is a plural means of referring to individuals. Second-order logic can also be applied to set theory. In this context we can interpret monadic second-order quantification over sets as plural quantification.

**See also** Computability Theory; First-Order Logic; Gödel, Kurt; Leibniz, Gottfried Wilhelm; Logic, History of; Modern Logic: From Frege to Gödel; Mathematics, Foundations of; Proof Theory; Quine, Willard Van Orman.

### Bibliography

- Boolos, George. *Logic, Logic and Logic*. Cambridge, MA: Harvard University Press, 1998.
- Boolos, George. “Nominalist Platonism.” *Philosophical Review* 94 (1985): 327–344. Reprinted in Boolos 1998.
- Boolos, George. “On Second-Order Logic.” *Journal of Philosophy* 72 (1975): 509–527. Reprinted in Boolos 1998.
- Boolos, George. “To Be Is to Be a Value of a Variable (or to Be Some Values of Some Variables).” *Journal of Philosophy* 81 (1984): 430–449. Reprinted in Boolos 1998.
- Enderton, Herbert. *A Mathematical Introduction to Logic*. San Diego, CA: Harcourt/Academic Press, 2001.
- Shapiro, Stewart. *Foundations Without Foundationalism: A Case for Second-Order Logic*. Oxford: Oxford University Press, 1991.
- Shapiro, Stewart. “Second-Order Logic.” In *Blackwell Guide to Philosophical Logic*, edited by Lou Goble. Oxford: Blackwell, 2001.
- Simpson, Stephen G. *Subsystems of Second-Order Arithmetic*. Berlin: Springer, 1998.
- Quine, W. V. O. *Philosophy of Logic*. Cambridge, MA: Harvard University Press, 1970.
- van Dalen, Dirk. *Logic and Structure*. Berlin: Springer-Verlag, 1994.

Jeffrey Ketland (2005)

### SELF

In its normal use the English expression “self” is not even quite a word, but something that makes an ordinary object pronoun into a reflexive one (e.g., *her* into *herself*). The reflexive pronoun is used when the object of an action or attitude is the same as the subject of that action or attitude. If I say Mark Twain shot *himself* in the foot, I describe Mark Twain not only as the shooter but as the person shot. In this sense “the self” is just the person doing the action or holding the attitude that is somehow in question. “Self” is also used as a prefix for names of



activities and attitudes, identifying the special case where the object is the same as the agent: self-love, self-hatred, self-abuse, self-promotion, self-knowledge.

“The self” often means more than this, however. In psychology it is often used for that set of attributes that a person attaches to himself or herself most firmly, the attributes that the person finds it difficult or impossible to imagine himself or herself without. The term *identity* is also used in this sense. Typically, one’s sex is a part of one’s self or one’s identity; one’s profession or nationality may or may not be.

In philosophy the self is the agent, the knower and the ultimate locus of personal identity. If the thought of future reward or punishment is to encourage or deter me from some course of action, I must be thinking of the person rewarded as me, as myself, as the same person who is now going to endure the hardship of righteousness or pass up the enjoyments of sin in favor of this ultimate reward. But this same self comes up in much more mundane transactions. If I pick up the cake and shove it in this mouth rather than that one, is it not because I think it will be me, the very same person who picks up the cake, that will have the pleasure of tasting it?

A straightforward view of the self would be that the self is just the person and that a person is a physical system. This view has been challenged on two fronts. First, the nature of freedom and consciousness has convinced many philosophers that there is a fundamentally non-physical aspect of persons. The second challenge stems from puzzling aspects of self-knowledge. The knowledge we have of ourselves seems very unlike the knowledge we have of other objects in several ways, and this has led some philosophers to rather startling conclusions about the self. In his *Tractatus*, Ludwig Wittgenstein tells us that “I am my world” and that “the world is my world” (1961, 5.63, 5.641). This should lead us to the rather surprising conclusion that I am the world, or that at least Wittgenstein was. He draws at least one conclusion that would follow from this: “at death the world does not alter, but comes to an end.”

The contemporary philosopher Thomas Nagel has been led to a possibly less radical but still quite dramatic view. According to Nagel, when he says “I am Tom Nagel,” at least in certain philosophical moods, the “I” refers to the “objective self,” which is not identical with but merely contingently related to the person Tom Nagel. This self could just as well view the world from the perspective of someone other than him (Nagel, 1983). We need to discuss the puzzling features of self-knowledge that give rise to such views.

## SELF-KNOWLEDGE

“Self-knowledge” seems to have a straightforward meaning: cases of knowledge in which the knower and the known are identical. But this does not seem sufficient. The philosopher Ernst Mach once got on the end of a bus and saw a scruffy, unkempt, bookish-looking sort of person at the other end. He thought to himself,

- (1) That man is a shabby pedagogue.

In fact, Mach was seeing himself in a large mirror at the far end of the bus. He eventually realized this and thought to himself:

- (2) I am that man.
- (3) I am a shabby pedagogue.

Now consider Mach at the earlier time. Did Mach have self-knowledge? In our straightforward sense it seems that he did. He knew that a certain person was a shabby pedagogue and, furthermore, that person was him. The knower and the known were the same. But this is not what we mean by self-knowledge. Self-knowledge is something Mach really had only when he got to step (3), when he would have used the word *I* to express what he knew.

Self-knowledge seems peculiar. First, it seems “essentially indexical.” Statement (3) expresses self-knowledge because of the word *I*; it is hard to see how Mach could have expressed self-knowledge without using the first person. If he said “Mach is a shabby pedagogue,” he would be claiming to know only what everyone else may have known. It does not seem that there is any objective characterization *D* of Mach, such that knowing that *he* is a shabby pedagogue amounts to knowing that *D* is a shabby pedagogue (Castañeda, 1966, 1968; Perry, 1990, 1993).

Secondly, we seem immune to certain sorts of misidentification with respect to self-knowledge. If we learn, in certain ways, that someone is in pain, then we cannot miss the fact that it is we who are in pain. That is, if Mach discovers that he has a headache in the ordinary way that a person discovers she has a headache, he can scarcely be wrong about *who* has the headache, if the range of choices is “I/you/that man,” and so forth. Of course he can be wrong if the range of choices is “Mach/Freud/Wittgenstein,” and so on, for he might not realize which of those people he is (Shoemaker, 1984).

Third, self-knowledge seems to play a unique cognitive role. If Mach desires that *he* do so and so, and believes that *he* can do so and so by executing such and such a

movement, then he will execute that movement without further ado (Perry, 1990).

### AGENT-RELATIVE KNOWLEDGE

At least some of these peculiarities of self-knowledge can be explained by taking self-knowledge to be a species of agent-relative knowledge. There are two quite different ways of cognizing objects (people, things, places, and times). We can think of them via their relationship to us, the role they are playing in our lives at the moment of thought: the object I see; the present moment; the place I'm at; the person I'm talking to. We need to think about things in the first way, when we are picking up information about them perceptually or interacting with them, since ways of knowing and acting are tied to these agent-relative roles. I can learn about the here and now by looking; I can learn about the person I am talking to by asking questions, and so forth.

But these agent-relative roles cannot be our only ways of thinking about objects of more than passing interest to us. Different objects play the same agent-relative roles at different times, and at any given time many of the objects we wish to retain information about will not be playing any agent-relative role for us. And we cannot accumulate information along such roles. Suppose I am in Tokyo on Tuesday but return to Palo Alto on Friday. From the facts that on Tuesday I truly thought "Japanese is the official language *here*" and on Friday I truly thought "Senator Stanford used to live near *here*" it does not follow that there is some place where Japanese is the official language and near which Senator Stanford used to live.

In order to retain and accumulate information about objects, to construct and maintain a coherent picture of the world, we need to have a way of conceiving of objects as existing independently of us, as occupying and then ceasing to occupy various agent-relative roles. That is, we need objective ways of thinking about objects. We keep track of them by names or descriptions that do not depend on their relationship to us: Cordura Hall, 4 p.m., June 23, 1995, the southernmost town in Santa Clara County, Aurora Fischer. These serve as our fundamental ways of thinking about those objects. Recognition consists in connecting our objective ways of thinking of objects with the roles those objects play at a given moment. Consider the knowledge I might express with "Today is July 4." This is knowledge that a certain day, objectively conceived ("July 4"), is playing a certain role in my life; it is the present day, the day on which the thinking and speaking take place. This kind of knowl-

edge, "knowing what day it is," is quite crucial to successful application of other, more objective knowledge. If I know that the party is on July 4 and know that today is July 4, then I will form the right expectations about what the day will be like.

Similarly, I may be in Kansas City and know that Kansas City is a good place for a steak dinner. But if I do not know that I am in Kansas City, if I do not realize that Kansas City is playing the "here" or "this city" role in my life at this moment, I will not be able to apply the knowledge that Kansas City is a good place for a steak dinner.

And again, I may know that Aurora Fischer has important information about my schedule, but unless I realize that the person I am talking to is Aurora Fischer, I will not apply this information and say, "Can *you* tell me where this afternoon's meeting is?"

These kinds of knowledge are, like self-knowledge, "essentially indexical." We use *now* and *today* to express our knowledge of what time it is and *here* to express our knowledge of where we are. These locutions are not reducible to names or objective descriptions, just as *I* was not. I cannot express what I say when I say, "The meeting starts right now" by saying "the meeting starts at *D*" for any description *D* of the present moment.

We are also immune to certain sorts of misidentification when we use certain methods of knowing. There is a way of finding out what is going on around one, namely opening one's eyes and looking (Evans, 1985). Now when one learns what is going on in this way, one can hardly fail to identify the time at which this is happening as now and the place as here. And finally, the forms of thought we express with *now* and *here* seem to have a unique motivational role. If I want to do something here and now, I will simply do it.

### SELF-KNOWLEDGE AS AGENT-RELATIVE KNOWLEDGE

"Self" is really the name of such an agent-relative role, that of identity. As with other agent-relative roles, there are special ways of knowing and acting that are associated with identity. If Mach had wished to know, during the interval while he was confused, if the shabby pedagogue he was seeing had lint on his vest, he would have had to walk over to him and look. If Mach had wanted to know if he himself had lint on his vest, he could have simply lowered his head and looked. Had he done this, he would have had no doubt about whom the lint was on. If Mach found lint and wanted to brush it off, he would engage in self-brushing, a quick movement of the hand across one's

front that each of us can use to remove lint from our own vest and no one else's.

Unlike the other agent-relative roles, identity is permanent. I will talk to many people, be in many places, live through many times in the course of my life. But there is only one person I will ever be identical with, myself. Hence, accumulation along "I" is valid, unlike accumulation along "here" or "now" or "that man."

Earlier we rejected the straightforward account of self-knowledge, as knowledge about a person by that very person. Now we can put forward an alternative. Self-knowledge is knowledge about a person by that very person, with the additional requirement that the person be cognized via the agent-relative role of identity. This agent-relative role is tied to normally self-informative methods of knowing and normally self-effecting ways of acting. When these methods are employed, there will be immunity of misidentification as to who is known about, or who is acted upon.

This role can serve as a person's fundamental concept of himself or herself. In this way our self-conceptions have structures that are different from our conceptions of other individuals of importance to us. If we understand the special way in which a person's self-knowledge is structured, we do not need to postulate anything but the person himself or herself for the knowledge to be about.

**See also** Consciousness; Freedom; Identity; Indexicals; Mach, Ernst; Nagel, Thomas; Personal Identity; Philosophy of Mind; Reduction; Self-Knowledge; Wittgenstein, Ludwig Josef Johann.

### **Bibliography**

- Barnes, A. *Seeing through Self-Deception*. Cambridge, U.K.: Cambridge University Press, 1997.
- Bermúdez, J. *The Paradox of Self-Consciousness*. Cambridge, MA: MIT Press, 1998.
- Brewer, B. "Bodily Awareness and the Self." In *The Body and the Self*, edited by J. Bermúdez, A. Marcel, and N. Eilan. Cambridge, MA: MIT Press, 1995.
- Cassam, Q. *Self and World*. Oxford: Oxford University Press, 1997.
- Cassam, Q., ed. *Self-Knowledge*. Oxford: Oxford University Press, 1994.
- Castañeda, H.-N. "'He': A Study in the Logic of Self-Consciousness." *Ration* 8 (1966): 130–157.
- Castañeda, H.-N. "On the Logic of Attributions of Self-Knowledge to Others." *Journal of Philosophy* 65 (1968): 439–456.
- Evans, G. "Understanding Demonstratives." In *Collected Papers*. Oxford: Clarendon Press, 1985.
- Evans, G. *The Varieties of Reference*. Oxford: Clarendon Press, 1982.

- Gallois, A. *The Mind Within, The World Without*. Cambridge, U.K.: Cambridge University Press, 1996.
- Gertler, B., ed. *Privileged Access: Philosophical Accounts of Self-Knowledge*. Aldershot, U.K.: Ashgate, 2003.
- Hill, C., ed. *Philosophical Topics: Introspection* 28 (2000).
- Lowe, E. J. *Subjects of Experience*. Cambridge, U.K.: Cambridge University Press, 1996.
- Moran, R. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton, NJ: Princeton University Press, 2001.
- Nagel, T. "The Objective Self." In *Knowledge and Mind*, edited by C. Ginet and S. Shoemaker. New York: Oxford University Press, 1983.
- Neisser, U., and D. Jopling, eds. *The Conceptual Self in Context: Culture, Experience, Self-Understanding*. Cambridge, U.K.: Cambridge University Press, 1997.
- Perry, J. *The Problem of the Essential Indexical*. New York: Oxford University Press, 1993.
- Perry, J. "Self-Notions." *Logos* 11 (1990): 17–31.
- Shoemaker, S. *The First-Person Perspective and Other Essays*. Cambridge, U.K.: Cambridge University Press, 1996.
- Shoemaker, S. *Identity, Cause, and Mind*. Cambridge, U.K.: Cambridge University Press, 1984.
- Strawson, G. "The Self." *Journal of Consciousness Studies* 4 (1997): 405–428.
- Wittgenstein, L. *Tractatus Logico-Philosophicus*. Translated by D. F. Pears and B. F. McGuinness. London: Routledge and Kegan Paul, 1961.
- Wright, C., B. Smith, and C. Macdonald, eds. *Knowing Our Own Minds*. Oxford: Clarendon Press, 1998.

**John Perry (1996)**

*Bibliography updated by Alyssa Ney (2005)*

## SELF-CONSCIOUSNESS

See *Consciousness*

## SELF-DECEPTION

If weakness of will is a pathology of agency, then it is natural to regard self-deception as a pathology of cognition. Self-deception is a species of motivated believing in which the cognition of a subject is driven by desire towards the embrace of some proposition—typically, "in the teeth of the evidence." Here we may think of the alcoholic, the terminal cancer patient, or the anorexic, who, even while in possession of compelling evidence of his condition, insists, sincerely, that it is just not so. Many investigators require that, more than this, the self-deceiver must be understood to bring about his deception intentionally and knowingly in pursuit of the doxastic embrace of some motivationally or affectively favored proposition. Were this so, self-deception would seem to involve the sort of deep or internal irrationality distinc-

tive of weakness of will. For just as the weak-willed individual knowingly and intentionally acts against her judgment of what she takes herself to have best or sufficient reason to do, so the self-deceiver, on this picture of the phenomenon, knowingly violates her own norms or standards of reasoning—she comes to believe what she also believes there is insufficient reason to believe.

Producing a coherent account of how this is so has proved a vexing matter. Other investigators have argued that self-deception can be fully explicated without appeal to a subject's intentionally aiming to bring about her own deception against her current regard for the facts, and therefore without implicating this sort of deep irrationality. Notwithstanding these disputes, it seems clear that when we charge a subject with self-deception, we aim to offer both an explanation of how it is that a subject came to hold or retain a belief and a negative appraisal of the subject's belief-forming behavior.

In a quite literal way, the impetus behind the philosophical problem of self-deception springs from the force and puzzlement attached, in certain circumstances, to the question "How *could* he believe that?" We are all familiar with various unpleasant features of our cognitive lives, and there is no doubt that we do reason in ways that, as a matter of fact, violate epistemic norms that we endorse (the term 'epistemic,' meaning of or relating to knowledge, is derived from the Greek, "epistēmē"). The sources of such failures are many: we are subject to a profound confirmation bias, prone to be taken in by the vividness and salience of data (Nisbett and Ross 1980), forgetful and subject to fatigue, and so forth. Very plausibly, self-deception raises more pressing difficulties. In such cases, securing an answer to the question "How *could* he believe that?" compels us to reflect upon such issues as the nature of belief, doxastic agency, the unity of the self, and epistemic rationality and irrationality, among many others.

## THE PHENOMENON

As suggested above, much controversy surrounds the effort to characterize the process of self-deception, the nature of the phenomenon itself, and the sort of irrationality characteristic of the phenomenon. Notwithstanding this disagreement, clear instances of what we call "self-deception" come readily to mind. The stock and shopworn example of the husband who, even though in possession of compelling evidence of his wife's infidelity, nonetheless insists upon her faithfulness is a case in point. Our husband may generate richly ornamented stories the apparent aim of which is to explain away the, by our lights, dispositive evidence of the fact of his wife's

affairs. He may focus upon the occasions on which his wife has displayed great solicitousness and affection towards him, and he may well regard these data as clear and compelling evidence of her continued love for him. Moreover, he may subject evidence that strongly points towards his wife's infidelity to sustained and withering critical scrutiny, while precipitately embracing data indicative of her continued faithfulness. In short, our hapless husband repeatedly searches for reassuring evidence and probes various hypotheses in a sustained and continuing fashion in order to arrive at and then to retain the favored belief against various threats. Core cases of self-deception would, then, appear to involve a subject engaging in strategies the aim of which is the embrace of some proposition(s).

## TRADITIONALISM ABOUT SELF-DECEPTION

How are we to characterize and explain such behavior? An approach to such cognitive misadventures that we can term "traditionalism" aims to assimilate the dynamics of self-deception to those of interpersonal deception. As Mary Haight writes: "[I]f A deceives B, then for some proposition(s) *p*, A knows that *p*; and either A keeps or helps to keep B from knowing that *p*, or A makes or helps to make B believe that  $\sim p$ , or both" (1980, p.8). These lexical considerations (Mele 2001) for the traditionalist view, then, make it perfectly natural to characterize our self-deceived husband as knowing or believing that his wife is unfaithful and as aiming and ultimately succeeding in bringing it about that he comes to believe that she is, in fact, a loyal spouse. On such a model, the husband is not simply credulous, not merely stupid or epistemically careless; nor is he simply seduced by the salience or vividness of various data, or taken in by the confirmation bias. He is not, then, in the view of the traditionalist, merely a wishful thinker or believer. He aims at his own deception; he works hard to deceive himself. How else, we may ask ourselves, can he possibly believe that his wife is faithful? Why does he engage in such byzantine strategies, the apparent point of which is to avoid the implications of the evidence? Because he knows, or at the least strongly suspects, the truth—that she is unfaithful.

A typical traditionalist, then, will hold that our husband:

1. Believes that his wife is unfaithful (or believes that he ought rationally to believe that his wife is unfaithful).

2. Engages in intentional activity the aim of which is the acquisition of the belief that his wife is faithful.

3. Believes, at least for a time, both the belief adverted to in (1) and the belief adverted to in (2).

Donald Davidson, in an extremely influential essay titled “Deception and Division,” embraced these three conditions. As he puts it in a much-cited passage:

The acquisition of a belief will make for self-deception only under the following conditions: A has evidence on the basis of which he believes that *p* is more apt to be true than its negation; the thought that *p*, or the thought that he ought rationally to believe that *p*, motivates A to act in such a way as to cause himself to believe the negation of *p*. The action involved may be no more than an intentional turning away from the evidence in favor of *p*, or it may involve the active search for evidence against *p*. All that self-deception demands of the action is that the motive originate in a belief that *p* is true ... and that the action be performed with the intention of producing belief in the negation of *p*. Finally, and this is what makes self-deception a problem, the state that motivates the self-deception and the state it produces co-exist. (1985, p. 145)

It should be noted that Davidson’s rationale for a contradictory or inconsistent belief requirement for self-deception is not—as it is on some accounts of self-deception (Demos 1960)—that the self-deceiver literally *lies* to himself. Davidson takes, very plausibly, the project of lying to oneself to require a self-defeating intention. Rather, Davidson takes the philosophical problem of self-deception to be a matter of our being forced to come to grips with a continuing and synchronous irrational or inconsistent state—a state he takes to be distinctive of self-deception. Davidson characterizes self-deception as a condition brought about by my intentionally causing myself to believe against what I also believe and *continue* to believe to be the weight of the evidence. As a result, and not surprisingly, Davidson argues that the characterization of such a state requires the postulation of mental partitions, divisions in the mind.

**THE PUZZLES OF SELF-DECEPTION.** Still, whatever the attractions of such an account, it is difficult to fathom just how this sort of mental gymnastics can be carried off. Immanuel Kant was, for example, clearly puzzled by the looming difficulties here; as he put it in his *Metaphysical Principles of the Virtues*, “Since a second person is required when one intends to deceive, deceiving oneself

deliberately seems in itself to contain a contradiction” (cited in Darwall 1988, p. 411).

In a bit more detail, traditionalism has been taken by many to give rise to two difficulties. First, there is what Alfred Mele has termed the “static puzzle” (1987, 2001). The very state of mind of the self-deceiver might strike us as deeply puzzling. How can it be that the self-deceiver believes that *p* and also believes that not-*p*? There is no doubt, of course, that human beings often harbor inconsistent beliefs, where one of the beliefs is repressed or otherwise not currently or fully available to a subject’s awareness. What is harder to understand is a case in which both such beliefs are fully available to a subject.

Second, such an account makes for a strategic puzzle. Annette Barnes puts a version of the difficulty so: if I am to be self-deceived, I must “as deceived, be taken in by a strategy that, as deceiver I know to be deceitful” (1997, p. 18). The self-deceiver might well, as Davidson suggests, intentionally turn his attention away from evidence supportive of the threatening belief and seek out evidence of the favored belief with the aim of inducing in himself the latter. But if this plan is to succeed, it is not easy to see how the self-deceiver could fail to be wholly taken in by his ruse. That is, a condition of success of such a project would appear to be that the conviction that he ought rationally to believe the epistemically sanctioned proposition be exiled or come to be regarded as epistemically undermined *before* he can come to accept that his favored proposition is true. This is, however, very near to the sort of gambit recommended by Blaise Pascal in order to induce belief in the existence of God. There is no doubt that we can intentionally bring about conditions the result of which is that we come to believe what, at the time we brought about those conditions, we took ourselves to have no good reason to believe. This, however, does not appear to make for the deep and synchronous irrationality stalked by Davidson.

It should be noted that more modest traditionalists, while rejecting the contradictory belief requirement, have argued that the cognitive biasing in self-deception must be intentional (Talbot 1995), or that the self-deceiver need only actively avoid troubling recalcitrant evidence (Bach 1997).

Notwithstanding the difficulties to which traditionalism about self-deception has been alleged to be prey, its attractions and allure are clear. It works admirably to capture some very powerful vernacular (and philosophical) intuitions about the phenomenon. Traditionalism would sharply distinguish self-deception from putatively less puzzling phenomena such as wishful believing, for the

self-deceiver knowingly and actively brings about her deception, while the wishful believer is merely duped. Insofar as the self-deceiver succeeds in getting herself to believe what she also believes is not so, she would appear to be guilty of a profound form of epistemic irrationality. In addition, the sort of doxastic tension, instability, and fragility the traditionalist aims to describe has seemed to many the hallmark of self-deception. Lastly, insofar as the self-deceiver intentionally and knowingly brings about her deception, she is clearly blameworthy.

Predictably, perhaps, the modeling of self-deception upon interpersonal deception has tended to provoke three sorts of response. The first is outright skepticism about the phenomenon. As Mary Haight puts it: “[S]elf-deception is literally a paradox. Therefore it cannot happen” (1980, p. 73). The second response is a reconceptualization of self-deception as less a purely cognitive or doxastic affair and more an existential (or “actional”) matter. Herbert Fingarette’s pioneering work, *Self-Deception* (1969), is notable example of this tack. He writes of the self-deceiver that he “is one who is in some way engaged in the world but who disavows the engagement, who will not acknowledge it even to himself as his. That is, self-deception turns upon the personal identity one accepts rather than the beliefs one has” (p. 66). In this respect, Fingarette’s is a powerful development and reworking of themes from Jean-Paul Sartre’s famous discussion of “bad faith.” Finally, in the third response one can cleave to the interpersonal model in literal fashion but seek to avoid the difficulties via a very robust partitioning or homuncularist account of self-deception. This is David Pears’s account. He writes that cases of self-deception are to be explicated by appeal to a “subsystem” or homunculus that “is built up around the wish for the irrational belief [e.g. the husband’s belief that his wife is faithful]. Although it is a separate centre of agency within the whole person, it is, from its own point of view, entirely rational. It wants the main system to form the irrational belief, and is aware that it will not form it, if the [belief that there is no good reason to so believe] is allowed to intervene. So with perfect rationality it stops its intervention” (1984, p. 87; see also Pears 1986). Mark Johnston (1988) develops a series of powerful objections (e.g., “Why should the deceiving subsystem be interested in the deception” (p. 64)) to homuncular explanations of self-deception.

## DEFLATIONIST ACCOUNTS OF SELF-DECEPTION

A second family of accounts, “deflationism,” aims to circumvent many of the difficulties the traditionalist regards

as fundamental to the posing of the problem of self-deception. Alfred Mele, Mark Johnston, and Annette Barnes have all developed noteworthy deflationist accounts. According to deflationists, self-deception is a matter of coming to believe that *p* as a consequence of biased cognitive processing that is itself the product of the various motivational states of the subject. Such accounts very often take their cue from a rejection of the lexical considerations in favor of traditionalism (Mele 1987, 1997, 2001; Barnes 1997; Johnston 1998). So, for example, it is plausibly argued that there are many clear cases of interpersonal deception that involve neither the deceiver’s knowledge of the proposition the deceived comes to believe, nor intentional deception. But if this is so, there is no obvious reason to require these conditions when it comes to the characterization and, ultimately, the explanation of self-deception. Rather, if, for example, the process of self-deceiving oneself must be understood to be mediated by the subject’s intention to come to believe the favored and epistemically suspect proposition, this must be established by appealing to the fact that an explanation of particular features of the phenomenon itself *requires* such intentional activity. This is what deflationists deny. Core cases of self-deception, it is insisted, can be fully explained without appeal to the psychological exoticism characteristic of many versions of traditionalism.

Alfred Mele’s is the most influential of deflationist accounts. According to Mele, the following conditions are jointly sufficient for a subject’s entering self-deception in acquiring a belief that *p*.

1. The belief that *p* which *S* acquires is false.
2. *S* treats data relevant, or at least seemingly relevant, to the truth value of *p* in a motivationally biased way.
3. This biased treatment is a nondeviant case of *S*’s acquiring the belief that *p*.
4. The body of data possessed by *S* at the time provides greater warrant for not-*p* than for *p*.

(2001, p. 51; see also Mele 1987, p. 127)

The account is notable for what it does not include. There is no requirement that the subject must intentionally bring about his deception, nor is there a contradictory belief requirement. It should be noted, as well, that the motivational states mentioned in (2) will typically be desires for states of affairs; for example, our husband’s desire that his wife be faithful. This is to be distinguished from familiar traditionalist accounts according to which our husband not only desires that his wife be faithful but,

in addition, desires that he believe (or come to believe) that his wife is faithful; it is by virtue of the possession of this latter desire that, by the lights of the traditionalist, the husband comes to self-deceive himself. (Dana Nelkin [2002] has argued that, on pain of counting cases that do not involve self-deception as self-deception, the deflationist, like the traditionalist, must appeal to a subject's desire to believe.)

Mele, in particular, has emphasized the ways in which the motivational states of a subject can harness various sources of cognitive bias. Our husband's desire that his wife is faithful may trigger positive misinterpretation of data, negative misinterpretation of data, and selective evidence gathering and attention. Moreover, familiar "cold" or unmotivated sources of bias may also be triggered by motivation. That our husband desperately wants his wife to be loyal may make data indicative of her faithfulness more vivid as well as more salient. (We do, after all, tend to think about the objects of our desires.) Additionally, it seems clear that motivation will influence the selection of which hypotheses we begin testing with and so may trigger the confirmation bias.

**DIFFICULTIES FOR DEFLATIONISM.** Needless to say, it has been argued that various features of core cases of self-deception render the deflationist account implausible. William Talbott (1995), for example, has argued that not only is intentional self-deception possible in a single coherent self but, additionally, that we must appeal to an agent's intention to bias her cognition in favor of a particular proposition regardless of the truth of that proposition, if we are to explain various distinctive features of the phenomenon.

First, the process of self-deception might be regarded as too complex, too light-fingered and strategic to be the result of a non-intentional mechanism or process. Indeed, in core cases of self-deception—cases like that of our husband—the subject explains away just what needs to be explained away, he searches for just the evidence he needs in order to come to believe the favored proposition, he does not look just where he must not look, and so forth. This is just the sort of behavior characteristic of means-end rationality, and so of intentional behavior. Moreover, if the processes mediating self-deception are nonintentional, if such processes are "launched" as a simple result of our inhabiting various motivational states, why is it that human beings do not invariably bias their cognition in the direction of motivationally favored propositions? Happily, we do not always become self-deceived that *p* when we powerfully desire that *p*. Self-

deception is in this sense "selective." Again, it would seem that an extremely plausible explanation of why it is that I do come to bias my cognition when I do is that I intend to do so. (It is to be emphasized that Talbott takes our self-deceptive intentions to bias our cognition to be *unconscious* intentions. Annette Barnes [1997] and Ariela Lazar [1999] have developed a number of powerful objections to the notion that unconscious intentions play a crucial role in the explanation of self-deception.)

**FACING THE QUESTION: "P OR NOT-P?"** Does the deflationist have the resources to respond to these difficulties? Much recent discussion of these issues has drawn on the social psychological investigation of lay-hypothesis testing. Consider the task of any hypothesis tester—including the prospective self-deceiver. He faces questions of the form: "*p* or not-*p*?" The effort to settle any such question will involve costs to the agent in the form of time, and energy spent in the task of hypothesis testing. What is central to this "pragmatic" account of hypothesis testing is another sort of cost involved in the settling of such questions: the cost of anticipated errors as noted by Friedrich (1993) and Trope and Liberman (1996). In aiming to settle a question, a subject aims to end her uncertainty, to reach her "confidence threshold" at which time hypothesis testing is ended. As such, there will be costs associated with settling the question in favor of *p*, when *p* is false (false positives), and costs associated with settling the question in favor of not-*p*, when *p* is true (false negatives). In brief, what is crucial to this account is that with regard to many such questions, the costs associated with such errors will be asymmetric rather than symmetric. As such, there will be what James Friedrich calls a "primary error," an error that the subject is preponderantly motivated to avoid. This error, not surprisingly, is fixed by the values, aims, and interests of the cognizer. Such asymmetric error costs, in turn, fix asymmetric confidence thresholds. The result is biased hypothesis testing and the striking appearance of intentional guidance toward the doxastic embrace of a favored proposition. As Friedrich (1993) puts it: "Lay hypothesis testers are always motivated by accuracy, in the sense that they want to detect and minimize particularly costly errors" (p. 357).

Consider the case of our husband. He must settle the question, "Does she or doesn't she?" His primary error is fixed by his desires and interests. As such, we can easily imagine that his primary error, the error he is most powerfully motivated to avoid, is the error of believing that his wife is unfaithful when she is not. This, then, generates asymmetric confidence thresholds. As a result, he will demand powerful and compelling evidence if he is to

accept that she is unfaithful, while requiring relatively little data to accept that she is faithful. As this is so, the model predicts that our husband will subject data suggestive of her infidelity to powerful critical scrutiny whereas he accepts data suggestive of her fidelity without serious investigation. The account promises a nonintentional explanation of the apparently strategic behavior of core cases of self-deception. It should not be forgotten, of course, that hypothesis testing is typically an amalgam of the intentional and non-intentional. *Any* hypothesis tester who faces the question “p or not-p?” does aim to settle that question. She knows, as well, of the means of which she must avail herself (seeking evidence, asking questions of those “in the know,” etc.) if she is to resolve her uncertainty. So the issue, it seems, is not whether the self-deceiver engages in any intentional behavior in coming to believe as she does. Rather, the issue is whether she *must* be understood to possess an intention to settle her question in some particular direction.

Moreover, it seems that the pragmatic account of hypothesis testing offers an explication of why it is that we do not invariably come self-deceptively to bias our cognition in favor of what it is that we anxiously desire to be so and, so, promises at least a tentative response to the selectivity problem. Again, whether an individual engages in biased hypothesis testing will be determined by the full range of the subject’s interests. So, for example, Talbott (1995) notes that hurtling down a steep mountain road and hearing unfamiliar and frightening noises when I depress my car’s brakes, I am not likely to come to believe that there is nothing amiss, even though there is no doubt that I very much want it to be the case that my brakes are just fine. Indeed, given that the error costs associated with believing my brakes are in working order when they are not are terrifically vivid, I may be likely to come to believe in biased fashion that my brakes are failing. (For skepticism concerning whether a pragmatic account of hypothesis testing holds an answer to the selectivity problem in its full generality see Jose Bermudez [2000].)

This last example indirectly raises the problem of “twisted” or “unwelcome” cases of self-deception (Mele 1999, 2001; Barnes 1997; Lazar 1999; Scott-Kakures 2000). It is indeed a striking fact that self-deception is not always a matter of coming—in biased fashion—to believe just what is desired (directly or indirectly) to be so. Indeed, overprotective parents come in strikingly biased ways to believe that their children are suffering from grave illnesses. Some subjects come to believe, on the basis of scant evidence, that their spouses are *unfaithful*. And, of course, we all have our favorite hypochondriac.

Though the matter is much disputed, such cases would appear to constitute at least a presumptive difficulty for familiar accounts of self-deception. Such cases do, however, appear to be explicable by appeal to the pragmatic account of hypothesis testing. Consider: A busy executive, driving to her work, is nearly hit by a careless motorist as she nears her freeway on ramp. As a result, it may be that she comes, later in her commute, to conclude that many drivers she passes are careless and so constitute a danger. This is not surprising, as she has been made vividly aware of the very high cost of failing to conclude that x is a bad driver if he is. As a result of these asymmetric error costs and the associated asymmetric confidence thresholds, she is apt to demand overwhelming evidence before concluding that x is a safe driver, and she is likely to require very little evidence to bring her to the conclusion that x is a bad driver.

According to the deflationist, then, the irrationality present in self-deception is not an irrationality that requires us to appeal to the traditionalist’s psychological machinery. Indeed, the irrationality present in self-deception is an irrationality with which we are all very familiar—it is a matter of biased reasoning. In this sense, self-deception, according to the deflationist, is not the cognitive pathology it has historically been understood to be. Much of the appeal to traditionalism springs from the intuition that only some distinctive cognitive pathology could explain the self-deceiver’s turning away from the proper aim of belief: truth. In this way, it may well be that, for the deflationist, the price of making self-deception appear more familiar is that what we are apt to regard as “normal” hypothesis testing will come to seem more suspect and less familiar.

*See also* Weakness of the Will.

### **Bibliography**

- Bach, Kent. “Thinking and Believing in Self-Deception.” *Behavioral and Brain Sciences* 20 (1997): 105.
- Barnes, Annette. *Seeing Through Self-Deception*. Cambridge, U.K.: Cambridge University Press, 1997.
- Bermudez, Jose. “Self-Deception, Intentions and Contradictory Beliefs.” *Analysis* 60 (2000): 309–319.
- Darwall, Stephen. “Self-Deception, Autonomy, and Moral Constitution.” In *Perspectives on Self-Deception*, edited by Amélie Oksenburg Rorty and Brian McLaughlin. Berkeley: University of California Press, 1988.
- Davidson, Donald. “Deception and Division.” In *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, edited by Ernest LePore and Brian McLaughlin. Oxford: Basil Blackwell, 1985. Reprinted in his *Problems of Rationality*. Oxford: Oxford University Press, 2004.



- Davidson, Donald. "Paradoxes of Irrationality." In *Philosophical Essays on Freud*, edited by Richard Wollheim and James Hopkins. Cambridge, U.K.: Cambridge University Press, 1982. Reprinted in his *Problems of Rationality*. Oxford: Oxford University Press, 2004.
- Demos, Raphael. "Lying to Oneself." *Journal of Philosophy* 57 (1960): 588–595.
- Fingarette, Herbert. *Self-Deception*. Berkeley: University of California Press, 1969. Reprint 2000.
- Friedrich, James. "Primary Error Detection and Minimization (PEDMIN) Strategies in Social Cognition: A Reinterpretation of Confirmation Bias Phenomena." *Psychological Review* 100 (1993): 298–319.
- Haight, Mary. *A Study of Self-Deception*. Atlantic Highlands, NJ: Humanities Press, 1980.
- Holton, Richard. "What Is the Role of the Self in Self-Deception?" *Proceedings of the Aristotelian Society* 101 (2001): 53–69.
- Johnston, Mark. "Self-Deception and the Nature of Mind." In *Perspectives on Self-Deception*, edited by Amélie Oksenburg Rorty and Brian McLaughlin. Berkeley: University of California Press, 1988.
- Lazar, Ariela. "Deceiving Oneself or Self-Deceived? On the Formation of Beliefs 'Under the Influence.'" *Mind* 108 (1999): 265–290.
- Mele, Alfred R. *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control*. New York: Oxford University Press, 1987.
- Mele, Alfred R. "Real Self-Deception." *Behavioral and Brain Sciences* 20 (1997): 90–102.
- Mele, Alfred R. *Self-Deception Unmasked*. Princeton, NJ: Princeton University Press, 2001.
- Nelkin, Dana. "Self-Deception, Motivation, and the Desire to Believe." *Pacific Philosophical Quarterly* 83 (2002): 384–406.
- Nisbett, Richard, and Lee Ross. *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, NJ: Prentice Hall, 1980.
- Noordhof, Paul. "Self-Deception, Interpretation and Consciousness." *Philosophy and Phenomenological Research* 67 (2003): 75–100.
- Pears, David. "The Goals and Strategies of Self-Deception." In *The Multiple Self*, edited by Jon Elster. Cambridge, U.K.: Cambridge University Press, 1986.
- Pears, David. *Motivated Irrationality*. Oxford: Clarendon Press, 1984.
- Sartre, Jean-Paul. *Being and Nothingness*. Translated by Hazel Barnes. New York: Philosophical Library, 1956.
- Scott-Kakures, Dion. "Motivated Believing: Wishful and Unwelcome." *Nous* 34 (2000): 348–375.
- Talbott, William. "Intentional Self-Deception in a Single Coherent Self." *Philosophy and Phenomenological Research* 55 (1995): 27–74.
- Trope, Yaacov, and Akiva Liberman. "Social Hypothesis Testing: Cognitive and Motivational Mechanisms." In *Social Psychology: Handbook of Basic Principles*, edited by E. Tory Higgins and Arie W. Kruglanski. New York: Guilford, 1996.

**Dion Scott-Kakures (2005)**

## SELF IN INDIAN PHILOSOPHY

The human phenomenological experience of the universe consists fundamentally of the self or subject encountering a world of objects. Thus the two main objects of philosophy are the subject or the self—its nature and constitution—on the one hand, and the universe, along with its nature and constitution, on the other. Indian philosophy is no exception to this rule.

This experiencing self is referred to by several terms in Indian philosophy, the one most widely used being *ātman*. The word is usually derived from the root *an*, which means "to breathe"; apparently the fact that the perceiving self is an animate being who faces other animate beings and inanimate objects is central to its emergence as the marker of the self. It is called *purua* when its distinction from inanimate nature or *prakti* is emphasized, and it is called *jīva* when the *ātman* is viewed as caught up in the cycle of *saṣāra* or birth and death, freedom from which becomes a goal of this empirical self (*jīva*). In many systems this freedom is attained when the *jīva* or empirical self discovers its true relationship to the *ātman* or metaphysical self. This is the essential theological structure of the school of Indian philosophy known as Vedānta. But virtually each school of Indian philosophy possesses its own conception of the self or *ātman*, which must now be examined. Such an examination is facilitated by a review of the conception of the self in each of the nine schools of Indian thought. Although this standardization is relatively recent (Halbfass 1988, p. 353) it is worth employing because it enables one to present the concept of the self across the various schools with some measure of coherence. These nine schools, usually listed in order, are the Cārvāka (of Lokāyata), Jaina, Bauddha, Nyāya, Vaiśeika, Sāṅkhya, Yoga, Mīmāsā, and Vedānta.

### CĀRVĀKA

According to the Cārvāka school, the body itself constitutes the self (*deha eva ātmā*); of course, what is meant is that the conscious body constitutes the self. However, this immaterial element of consciousness in the body is considered an epiphenomenon of the material components of the body, in a manner reminiscent of scientific materialism. The Cārvāka school would establish the plausibility of the emergence of a property not contained in the elements by their coming together on the analogy of water, which possesses the quality of wetness, a property not possessed by the two gases of which it is composed. There is no question then of postmortem survival

according to this school, as consciousness perishes with death. It therefore emphasizes making the most of life, with a pleasant death serving as the counterpart of salvation. Thomas McEvelley (2002) notes that these doctrines are similar to the ones Plato attributes to the *physiologoi*.

## JAINA

According to the Jaina school, the self consists of the soul or *jīva* which occupies the body. The soul is formless but it can occupy a body just as light might occupy a room. It is a striking feature of the Jaina view of the self that this *jīva* is said to be coextensive with the body. In view of the fact that the soul occupies the body, it can be said to occupy space, as the body does, and may even be said to be capable of extension, as when the body grows. The whole range of existence, including plants and minerals along with insects and so on, possesses a conscious soul, and if such consciousness—which is characteristic of the soul—is not apparent, it is because it is dormant under the influence of karma. In Jainism karma is considered a very fine material substance that can permeate a soul, just as motes of dust might permeate light. Jaina soteriology consists of ridding the *jīva* of such matter, which keeps it weighed down in *saṣāra*, so that, freed from it, it can rise to the top of the “universe” and be free forever. According to Jainism, knowledge is the natural attribute of the *ātman*, which is kept in check by *ajīva* or inanimate components of our being. “The eyes, for example are viewed here not as an aid to seeing, but as a check in the absolute sight of the soul” (Hiriyanna 1949, p. 61).

## BAUDDHA

While the Cārvāka school does not believe in an *ātman* and denies anything like liberation, and the Jaina school believes in both, Buddhism denies the existence of a self or *ātman* while upholding liberation from rebirth in the usual Indic sense. According to Buddhism, continuity is possible without identity; hence there is no need to postulate a self that is reborn, for the next birth can be viewed as being caused by the present in the process of coming to an end, like an echo. Nirvāṇa brings silence to the re-echoing chamber of *saṣāra*. The Buddhists seem to create many apparent logical difficulties for themselves by denying a permanent self or *ātman* but according to them the other systems create their own existential problems by believing in one. The Buddhist critique of a substantial ontology is very thoroughgoing; according to this critique, nothing whatsoever in this world possesses a permanent substratum (*sabbe dhammā anattā*). The permanence or lack of it in the self has been a major issue

in the Hindu-Buddhist interface (Chakrabarti 1999, chapter 5, appendix).

## NYĀYA AND VAIŚEIKA

The concepts of the self in the Nyāya and the Vaiśeika schools have much in common and hence are presented together. According to the Nyāya and the Vaiśeika school, the soul or *ātman* is eternal, but consciousness is not its inherent property. Consciousness arises when the self or *ātman* is conjoined with *manas* or the mind, which is, however, by itself inert. The soul or *ātman* differs from other atomic or all-pervasive objects in that, unlike them, it is potentially capable of consciousness. The selves are numerous and all-pervading but remain distinct in the state of release because of the property of *viśeṣa*, which accounts for things being different that are in other respects all alike—for example, two atoms that are otherwise identical are not numerically one. The self has no consciousness in the state of release because such a state involves the absence of *manas*. The *ātman* in Nyāya is a unique substance that possesses the attributes of cognition, emotion, and conation and the qualities of desire, aversion, pleasure, pain, volition, and knowledge. The Vaiśeika school provides a longer list (Organ). As these are not perceived by the external senses and are not physical, they must belong to a nonphysical substance such as the soul. However, although consciousness or knowledge is an attribute of the *ātman*, it is not inseparable from it. The soul is thus an independent substance, but consciousness is an accidental property of it. In order for conscious states to arise, *manas* must come into play, hence the otherwise cryptic remark that “the true self is broken up here, we may say, into two ‘selfless elements’” (Hiriyanna 1949, p. 91). Scholars such as McEvelley (2002) note parallels here with Aristotelian thought.

## SĀNKHYA AND YOGA

The concepts of the self in the Sāṅkhya and the Yoga schools are also sufficiently similar to be treated together. In Sāṅkhya the self is called *purua* or soul and represents pure consciousness, in opposition to *prakti*, which represents matter. The self loses its inherent consciousness by mistakenly identifying itself with the body as involved in the process of *saṣāra*; the self is utterly passive and merely a spectator but mistakes itself for an actor and thus undergoes the ups and downs of the cosmic drama. Although the word *purua* is often used in the singular, in reality the system allows for a plurality of *puruas*, all consisting of pure consciousness, but distinct from each other and *prakti* or matter. The *purua* in Sāṅkhya and Yoga

is an uncaused, eternal, all-pervading, and changeless reality, which witnesses change as a transcendent subject distinguished by pure consciousness that can itself never become an object of knowledge. Salvation consists of this discrimination (*viveka*) that one is pure spirit and not the mind with whose derivative reality one identifies oneself. The system of Yoga with its eight limbs or constituent elements is meant to guide one, through a series of meditations, to the realization of this ultimate transcendent witnessing subject as distinct from the mind, body, and ego just as the surface of the mirror is totally independent of the objects that are reflected in it but appear included in it.

### MĪMĀSĀ

The concept of the self in Mīmāsā is broadly similar to that found in Nyāya and Vaiśeika, but there are some differences. The list of specific qualities characterizing the self is similar but not identical, with Mīmāsā dropping those of dharma and adharma and adding that of *śakti* or potency. The most significant difference however consists of the fact that while according to the Nyāya-Vaiśeika school knowledge is a quality of the self, according to Mīmāsā it is an activity of the self.

### VEDĀNTA

The conception of the self or *ātman* in Vedānta needs to be presented in accordance with the school of Vedānta involved—whether it is Advaita Vedānta, Viśiādvaita Vedānta, or Dvaita Vedānta. Thus the exact conception of the *ātman* depends on whether we are dealing with the “non-dualism of the qualified” (Advaita) or dualism (Dvaita). Prior to identifying the self in these three schools of Vedānta, however, it might be useful to indicate the concept of the *jīva* they all share in common on the basis of their reliance on the same foundational texts. Another aspect of the issues relating to the self or *ātman*, which receives relatively greater treatment in Vedānta than in other systems, is its relationship to *Brahman*, or the ultimate reality. It will therefore be useful to begin the discussion of the self in the three Vedantic schools with the conception of it they all share and conclude it with their views on the nature of the relationship of this *ātman* to *Brahman*.

The description of the human person as found in the *Taittirīya Upaniṣad* (II, 1–5) became paradigmatic in later Vedānta. According to this description a person consists of five sheaths within which the *ātman* lies enclosed. Starting from the outside, the first sheath consists of the body made of food (*annamaya-kośa*); within it are the

vital airs that comprise the second sheath (*prāamaya-kośa*). The mind comprises the third sheath (*manomaya kośa*), consciousness the fourth (*vijñānamaya kośa*) and bliss the fifth (*ānandmaya*). In Advaita the self consists of self-effulgent consciousness (*svapṛakāsa caitanya*), which is rather than has consciousness. It is one and the same in all human subjects (unlike Sākhya) and eternally free. Later Vedānta also developed a doctrine of the three bodies that comprise a human being, which ostensibly seems to possess only one body. These are the *sthūla-śarīra* (or gross body) which corresponds to the *annamaya kośa*; the *sūkma-śarīra* (or subtle body), which corresponds to the *prāamaya*—the *manomaya*—and the *vijñānamaya kośa* and the *kāraa-sarīra* (or casual body), which corresponds to the *ānandmaya kośa*. The true self—the *ātman*—lies beyond all the five sheaths and the three bodies or may be said to constitute their nucleus depending on how one chooses to describe it (Kesarcodi-Watson 1994).

According to Advaita Vedānta, the *ātman* is one’s true self and is identical with *Brahman*. Any differences between the two are adventitious, caused by *upādhis* or superimpositions. A popular metaphor illustrates the point as follows: Different jars of different shapes and sizes may contain jar-space. The space enclosed by these jars may appear distinct, but if one breaks the jars, all space becomes one and the same. It was, however, one and the same to begin with—the jars only created ultimately artificial and unreal differences. Thus the selves of all are identical with each other and with *Brahman*.

The *ātman* per se is of the nature of pure consciousness according to Viśiādvaita Vedānta. The self is not pure consciousness, as maintained by Advaita Vedānta, but “a conscious subject called the ego or the ‘I.’” The *ātman* is the self, but this self both *is* and *has* consciousness. Moreover, the self may mistakenly identify with the objects of the world, but it is identical neither with them nor with *Brahman*. It has lost sight of its true nature, one of utter dependence on God or *Brahman*. *Moka* consists in being properly aligned with God through devotion and grace. One important difference between Advaita Vedānta and Viśiādvaita Vedānta is that whereas the *ātman* is infinite in its true nature according to Advaita Vedānta, it is considered atomic or infinitesimal in size in Viśiādvaita Vedānta, but it is able to have knowledge beyond itself through the fact that it not only is but possesses consciousness called *dharmabhūta jñāna*. *Jīvas* or empirical beings are infinite in number according to both the schools, but because of its metaphysical non-dualism, Advaita ultimately concedes only one reality: *ātman* = *Brahman*.

According to Dvaita Vedānta, the *ātman*s are infinite in number. The reason given to justify this is the obvious differences in their experiences (which are considered ultimately only empirical in Advaita Vedānta). They are atomic in size, and, as pointed out, differ from each other. They also differ from God, and the distance posited between them and God is somewhat greater in Dvaita Vedānta than in Viśiādvaita, as indicated by the very designations of these systems. Viśiādvaita Vedānta accepts the “monism of the qualified,” of God as qualified by the *ātman*s, but Dvaita Vedānta is frankly dualistic. Salvation results from the grace of God.

An utterance found in the Chāndogya Upaniṣad famously states “that thou art.” The *that* here is usually taken to relate to *Brahman* and the *thou* to *ātman*, and the interpretation of this seminal utterance in the three schools of Vedānta—the Advaita, the Viśiādvaita, and the Dvaita—is instructive of the differences in the concept of the *ātman* as it is understood in the three schools. According to Advaita Vedānta it means that *ātman* and *Brahman* are identical. “The identity of the denotation of the two terms” has to be realized “while their connotations are different” (Hiriyanna 1949, pp. 163–164). According to Viśiādvaita Vedānta it is to be interpreted as follows: “That’ finally denotes God as having the entire universe as his body; and ‘thou,’ God having the individual soul as his body” (p. 184). According to one interpretation offered by Dvaita Vedānta, the identity here really implies resemblance, for *ātman* “have features like sentience and bliss (though qualified) common with God” (p. 192). The precise idea of the self differs in virtually every system of Indian thought beyond the ones discussed here (see K. P. Sinha 1991).

**See also** Atomic Theory in Indian Philosophy; Brahman; God in Indian Philosophy; Knowledge in Indian Philosophy; Liberation in Indian Philosophy; Meditation in Indian Philosophy.

### Bibliography

- Chakrabarti, Kisor Kumar. *Classical Indian Philosophy of Mind: The Nyāya Dualist Tradition*. Albany: State University of New York Press, 1999.
- Chatterjee, Satischandra and Dhirendramohan Datta. *An Introduction to Indian Philosophy*. Calcutta: University of Calcutta, 1950.
- Halbfass, Wilhelm. *India and Europe: An Essay in Understanding*. Albany: State University of New York Press, 1988.
- Hiriyanna, Mysore. *The Essentials of Indian Philosophy*. London: George Allen & Unwin, 1949.

- Kesarcodi-Watson, Ian. *Approaches to Personhood in Indian Thought*, edited with a foreword by J. G. Arapura. Delhi: Sri Satguru Publications, 1994.
- McEvelley, Thomas. *The Shape of Ancient Thought: Comparative Studies of Greek and Indian Philosophers*. New York: Allworth Press, 2002.
- Organ, Troy Wilson. *The Self in Indian Philosophy*. The Hague: Mouton, 1964.
- Sinha, K. P. *The Self in Indian Philosophy*. Calcutta: Punthi Pustak, 1991.

Arvind Sharma (2005)

## SELF-INTEREST

Aristotle claims in the *Nicomachean Ethics* that it is the virtuous person “more than any other sort of person who seems to be a self-lover. ... he awards himself what is finest and best of all” (1168b28–30). Aristotle’s thought is that if one pursues things such as pleasure and wealth, one pursues what is base, injuring oneself. Contrast this with the implication of the recommendation “Look out for number one.” This advice is not taken to mean that one should pursue virtue. Rather, the idea is that the interests of others should take second place to one’s own. Virtue is not usually seen as the path of self-interest, especially because it can often involve self-sacrifice. This conflict suggests that effective pursuit of self-interest, or the interests of others, requires an account of the nature of well-being. (Henceforth, I will often use the term *well-being* rather than *self-interest* since that term is used more often in philosophical discussions of self-interest.) In the first part of this article, the major theories are discussed. In the second part, the focus is the importance (or lack thereof) of having an account of well-being for ethics.

### THEORIES OF WELL-BEING

The three dominant types of theory regarding well-being are hedonism, desire theory, and objective-list theories. This classification needs refinement, but it is a useful starting point. Take hedonism first. Jeremy Bentham (1776) was probably the most notorious proponent of hedonism. He espouses a type of hedonism that Derek Parfit dubs “narrow hedonism.” Bentham holds that pleasure is what is good for humans; pain is bad. He says, in *An Introduction to the Principles of Morals and Legislation*, that pleasures are homogenous sensations. How well one’s life is going depends on quantity of pleasure—the more the better.

One major objection to this outlook is that there is no felt sensation in common among the experiences that

people find pleasurable. For example, L.W. Sumner (1986) asks us to imagine the difference between the pleasure of going for a walk in the woods and the pleasure of completing a difficult task. Both are pleasures but they have no felt sensation in common.

Henry Sidgwick's form of hedonism, "preference hedonism," avoids this difficulty. He observes, in *The Methods of Ethics*, that "the only common quality [among pleasures] ... seems to be the relation to desire and volition expressed by the general term 'desirable'" (1907, p. 127). Sidgwick says that the judgments of the individual about which feelings are desirable must be taken as final. So pleasures, on this view, are those mental states that are desired by the individual. Some have noted that it strains the meaning of "pleasure" to call all of the mental states that we desire "pleasures." James Griffin's (1986) example citing Freud's desire to be mentally aware, but in horrible pain, rather than take opiates for his cancer pain, is such a case. Perhaps the name of the theory should be modified (as Shelly Kagan suggests) to "preference mental statism."

One strength of preference hedonism is that it respects the authority of the individual in determining which experiences make his or her life go better. Narrow hedonism says that a life of pleasurable sensation is better for the person even if one does not prefer it. Preference hedonism's weakness is that there are some desirable states of affairs that seem to contribute to well-being yet are not, strictly speaking, experiences. Probably, the most famous illustration of this problem is Robert Nozick's (1974) Experience Machine. Nozick asks us to imagine a machine that will give us all of the experiences we desire. He suggests that people would not choose to enter the machine because the experiences would have no relation to reality. Take another case. Imagine that someone is happy because she believes, falsely, that she has devoted friends. Now imagine the same person with happiness resulting from a true belief in devoted friends. Some think that the second is clearly the better life, especially if the second is preferred. The implication is that the fulfillment of desires for things other than mental states contributes to well-being. So, it seems that preference hedonism should be abandoned in favor of desire theory.

If desire theory is unrestricted, then it says that the fulfillment of any desire contributes to self-interest. That this is implausible is nicely shown by Parfit's (1984) case of the stranger: I meet a stranger with a supposedly incurable disease. I desire a cure for him; later, he is cured, though I never know this. It seems ludicrous to say that I am better off when the stranger is cured. This shows that

the desires that should count as contributing to a person's well-being have to be restricted. Parfit suggests that the desire has to be a desire about one's own life. This encompasses, for example, the desire not to be deceived, if it is a desire about one's own life.

But it may be unclear what qualifies as a desire about my own life. Is my desire to live in a just world, or in a world without starvation, a desire about my own life or not? Shelly Kagan (1992) argues that for a state of affairs to matter to my well-being, it has to affect me. My subjective experience is the same whether I am deceived or not. Kagan concludes that it may be that we should restrict the class of desires that are relevant to well-being to desires about mental states. This would mean a return to some form of preference hedonism. Whatever account is better, it is clear that a successful desire theory needs a plausible way to restrict the class of desires that impact well-being.

Now, consider the third type of theory: objective list. According to these theories certain things are good for people, even if they do not want them or have a negative attitude toward them. Consider John Rawls's (1999) famous example of the talented mathematician who wants to spend his life counting blades of grass. Some think that such a life cannot be good for him because the activity is worthless.

However, the difficulty for objective-list theories lies in giving an account of which activities are objectively worthwhile. One prominent account is Aristotle's Function Argument. The function of a flautist is to play the flute, and the flourishing flute player plays the flute well. The function of a human being is to engage in rational activity in accordance with virtue. A good example of a flute player plays the flute well, and a good example of a human engages in virtuous rational activity.

One major worry for the argument, noted by Peter Glassen (1957), is that even if it gives a correct account of human excellence, the inference that it must be good for a human to be a good example of his or her kind is fallacious. It is easy to imagine cases in which the excellent thing fails to be good for the agent. There are prosperous, sensible knaves, and sometimes the good die young. Some form of desire theory is now most commonly thought to be a correct account of well-being.

## WELL-BEING AND ETHICS

It may be thought that it is obvious why having a theory of self-interest is important for ethics. If moral theories yield principles about people's duties, and if their duties

include benefiting themselves and others, people need to know what counts as a benefit and what counts as a harm. The classical utilitarians—Jeremy Bentham, John Stuart Mill, and Henry Sidgwick—all think that the only thing that is intrinsically good is welfare or well-being and that the ultimate principle of morality is to perform the action that maximally benefits people.

Other theorists think that there are other goods besides well-being. W. D. Ross (1930) imagines two worlds in which there are equal amounts of happiness and equal amounts of virtue and vice. In the first the virtuous are happy, in the second the vicious are. Ross thinks that the first world is clearly better because of the distribution, even though they contain equal amounts of happiness. G. E. Moore holds, in *Principia Ethica* (1903), that it is good for beauty to exist even if it never affects anyone's conscious life. The deontologists writing in the Kantian tradition think that there is a duty of beneficence, although what is unconditionally good is the good will. However, there are some moral theorists who think that issues about well-being have little importance for ethics. For example, T. M. Scanlon (1998) argues both that individuals do not use the concept of well-being much in their deliberations about their own lives and that moral and political philosophers focus on just distributions of things such as primary goods, resources, or capabilities, rather than well-being. And he thinks that we do not have a general duty of beneficence. Notice, however, that one of the main reasons for the focus on primary goods or resources is the problem of expensive tastes. For example, Ronald Dworkin (1981) imagines a person who needs ancient claret and plover's eggs to be satisfied. Another person might reach an equal level of well-being with something much cheaper such as beer. To equalize welfare would require giving more resources to the first person. Dworkin and other theorists think that would be unjust, so they reject the idea of distributing welfare. The rejection might be correct, but it would be impossible to make the argument without a conception of human well-being.

**See also** Aristotle; Bentham, Jeremy; Dworkin, Ronald; Egoism and Altruism; Ethical Egoism; Eudaimonia; Freud, Sigmund; Happiness; Hedonism; Mill, John Stuart; Moore, George Edward; Nozick, Robert; Parfit, Derek; Pleasure; Rawls, John; Ross, William David; Sidgwick, Henry.

### **Bibliography**

Aristotle. *Nicomachean Ethics*. Translated by Terence Irwin. Indianapolis, IN: Hackett Publishing, 1985.

- Bentham, Jeremy. *An Introduction to the Principles of Morals and Legislation*, edited by J. H. Burns and H. L. A. Hart. London: Athlone Press, 1970.
- Dworkin, Ronald. "What is Equality? Part I: Equality of Welfare." *Philosophy and Public Affairs* 10 (1981): 185–246.
- Glassen, Peter. "A Fallacy in Aristotle's Argument About the Good." *Philosophical Quarterly* 7 (1957): 319–322.
- Griffin, James. *Well-Being: Its Meaning, Measurement, and Moral Importance*. Oxford: Clarendon Press, 1986.
- Kagan, Shelly. "The Limits of Well-Being." *Social Philosophy and Policy* 9 (1992): 169–189.
- Mill, John Stuart. "Utilitarianism." In *Essays on Religion, Ethics, and Society. The Collected Works of John Stuart Mill*. Vol. 10, edited by J. M. Robson. Toronto: University of Toronto Press, 1969.
- Moore, G. E. *Principia Ethica*. Cambridge, U.K.: Cambridge University Press, 1903.
- Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books, 1974.
- Parfit, Derek. *Reasons and Persons*. Oxford: Oxford University Press, 1984.
- Rawls, John. *A Theory of Justice*. 2nd ed. Cambridge, MA: The Belknap Press of Harvard University, 1999.
- Ross, W. D. *The Right and the Good*. Oxford: Oxford University Press, 1930.
- Scanlon, T. M. *What We Owe to Each Other*. Cambridge, MA: The Belknap Press of Harvard University, 1998.
- Sidgwick, Henry. *The Methods of Ethics*. 7th ed. London: Macmillan, 1907.
- Sumner, L. W. *Welfare, Happiness, and Ethics*. Oxford: Clarendon Press, 1986.

*Joyce L. Jenkins (2005)*

## SELF-KNOWLEDGE

Legend has it that when Chilan of Sparta asked, "What is best for man?" Apollo replied, "Know thyself." Thus, carved into the lintel of the Oracle of Apollo at Delphi were the Greek words "*gnothi seauton*"—"Know thyself" (Parke 1933). We can try to follow this Delphic injunction because we are self-conscious beings, capable of self-reflection.

Sigmund Freud (1923) maintained that we have unconscious beliefs, desires, motives, and intentions, and that extensive use of psychoanalytic techniques is often required to uncover them. Whether there is a Freudian unconscious is controversial, as is whether or not there is suppression or repression in the psychoanalytic senses. Nevertheless, our mental lives can be dissociated. And self-reflection can be as biased as reflection on any topic. Too charitable an attitude towards ourselves can leave us overly sanguine about the strength of our characters or the goodness of our intentions. Too uncharitable an attitude can lead to an exaggerated view of our frailties: We

may see ourselves as more selfish, less kind, and less well-intentioned than we really are. We can engage in wishful thinking, believing something about ourselves on less than adequate evidence because we want it to be true; evasive thinking, which involves turning our attention to other matters when thoughts about ourselves arise that conflict with our self-image; and skeptical thinking, in which we construct hypotheses on the fly to explain away evidence that conflicts with our self-image. Arguably, we can practice self-deception about our motives and reasons, but, in any event, we can certainly unintentionally mislead ourselves about them (McLaughlin and Rorty 1986).

Social psychologists have found that we have a tendency to confabulate. Asked to explain our decisions or actions, we sometimes fabricate explanations, not with an intent to deceive, but apparently with an eye toward justifying or making sense of those decisions or actions; the end result is we are taken in by our own fabrications. Evidence for this proclivity has led some philosophers to maintain that confabulation is so pervasive that our self-reports are just unreliable stories that we tell about ourselves for a variety of aims (Dennett 1991). Many philosophers, however, deny that our fallibility warrants skepticism about the possibility of self-knowledge. And many follow Socrates in holding that there is virtue in heeding the Delphic injunction. The quest for self-knowledge is a way of taking responsibility for ourselves. But it should, of course, be balanced with other activities of value. Yet another Delphic injunction is “Everything in moderation.” Narcissism or self-hatred can result in self-absorption, which is a vice.

Other people are better judges of certain aspects of our character than we are. They sometimes read our emotional state better than we do or remember our view about a certain topic better than we do. Moreover, they can tell us when we are confabulating or not being honest with ourselves. Nevertheless, it seems that each of us is able to know some things about ourselves in ways unavailable to others. Gilbert Ryle denied this claim, arguing that “the sorts of things that I can find out about myself are the same as the sorts of things that I can find out about other people, and the methods of finding them out are much the same. . . . John Doe’s ways of finding out about John Doe are the same as John Doe’s ways of finding out about Richard Doe” (1949, p. 155). He further claimed that “our knowledge of other people and ourselves depends on noticing how they and we behave” (1949, p. 181). According to behaviorism, we can know our own mental states only by observing our own behav-

ior or relying on the testimony of others who have. Of course, often others are better positioned to observe our behavior than we are. Hence, the joke “One behaviorist meeting another on the street said, ‘You feel fine! How do I feel?’” (Ziff 1958). The behaviorist view of self-knowledge seems untenable. We need not rely on observations of our behavior to know whether we are in pain, or are visualizing a red sunset, or are just now thinking to ourselves that behaviorism is untenable.

René Descartes (1985) drew attention to an area of mental life to which we seem to have first-person privileged access and with respect to which we seem authoritative: namely, our current conscious states. Conscious states include bodily sensations (aches, pains, itches, tickles, and the like), sense experiences (visual, auditory, and so on), mental imagery, felt emotions (feelings of fear, and so on), felt urges, and occurrent thoughts. We seem able to know our current conscious states in a way different from the way in which we know those of others. Indeed it seems that to know whether we are in a certain conscious state, we need only turn our attention to whether we are. To know whether we are in pain, for instance, it seems that we need only turn our attention to whether we are in pain.

Of course, we are by no means omniscient about such matters. Beliefs about what conscious states we are in involve the exercise of concepts (Sellars 1963, sec. 62); and we may lack the requisite concepts to know that we are in a conscious state of a certain sort. Even when we have the requisite concepts, we can fail to know simply because of lack of attention to the matter. Moreover, our concepts of types of conscious states are vague. Over the course of a morning we may move gradually from feeling cold to feeling warm, being unable to discern a difference in our thermal sensations from one moment to the next. On route to feeling warm we will pass through borderline cases of feeling cold; and in such cases we cannot know whether we feel cold. Theories of vagueness differ over why we cannot know in such borderline cases. According to semantic theories, the reason is that there is no fact of the matter whether the concept of feeling cold applies, and so they present no limitation to self-knowledge. But given our inability to discriminate cases falling very near the borderline from cases on it, our ability to know whether we feel cold may stop short of the borderline (Williamson 2000).

We can make verbal mistakes in our reports of our conscious states (Broad 1925), and even perhaps conceptual mistakes in our judgments about what conscious states we are in because of less than a full mastery of a rel-

evant concept (Burge 1979). But it has been held that if one has mastered the relevant concepts, one's belief that one is in a certain conscious state will be infallible and incorrigible (Ayer 1940). A belief or thought is infallible just in case it cannot be false; incorrigible just in case it cannot be shown to be false. Descartes (1985) argued, "*Cogito ergo sum*"—"I think, therefore I am"—taking his first-person thought that he thinks to be infallible, incorrigible, and indeed indubitable, such that it cannot be rationally doubted.

"Cogito-thoughts" such as that I am now thinking, and that I am now thinking that P, are indeed infallible, and hence incorrigible: they are true by virtue of my thinking them (Burge 1988). (Similarly, the belief that one has beliefs is true by virtue of one's having it.) Our infallibility in these cases, however, is not due to privileged access to the mental acts of thinking in question. If I write in English that I am writing in English, then what I write is true by virtue of my so writing it; even though I lack privileged access to whether I am writing in English, or even to whether I am in fact writing at all. Indeed, there are scenarios in which I am writing that I am writing in English but in which I fail to know that I am writing in English. The cogito-thought that one is thinking that P is (normally) an expression in consciousness of one's belief that one is thinking that P. But one can believe that one is thinking that P, when the only thought one is having is the thought that P; indeed that is the typical case (McLaughlin and Tye 1998a).

Beliefs to the effect that one is thinking that P are not true by virtue of one's having them. Moreover, they are fallible. To note just one reason: the longer it takes one to occurrently think that P, the more demand is put on short-term memory, and so the less reliable is one's belief that one is thinking that P (Armstrong 1963). Even the belief that we are in pain is fallible. Someone mesmerized by his guru might mistakenly believe that he is in pain solely on the basis of his guru's testimony to that effect. To take a more mundane case, upon hearing the start of the dentist's drill, one might momentarily mistake a feeling of pressure for a feeling of pain (Goldman 2002). (See also the "fraternity initiation" case described in Hill 1991, pp. 128–129.)

The term *introspection* is sometimes used very broadly to cover nearly any first-person, nonconsciously inferential avenue to knowledge of what mental states we are in. But on a more restricted usage (one to be followed here), introspecting a mental state is supposed to be a kind of direct act of awareness of the state. According to introspectionism, we can attend to our current conscious

states by introspecting them (Locke 1690; Broad 1925; Armstrong 1963; Hill 1991; Lycan 1996; Macdonald 1998, 1999; McLaughlin 2000, 2001, 2003c; Sturgeon 2000; Goldman 2002). The term *introspection* derives from the Latin *spicere*, which means "look," and the Latin *intra*, which means "within." But the etymology is misleading. Introspectionists do not hold that we literally look within. There is no "mind's eye" by which we observe our visual experiences, no "mind's ear" or "mind's toe" by which we observe, respectively, our auditory experiences and tactile experiences.

It is widely held that we see the scenes before our eyes by having visual experiences caused by them. We are not, however, aware of our visual experiences by having visual experiences caused by them. We do not see our visual experiences; they do not look any way to us. (Nor do they look any way to an internal homunculus; an untenable view that leads to an infinite regress of sighted homunculi embedded within sighted homunculi.) Introspective access is direct in a way perceptual access is not. We experience our experiences, not by having experiences of them, but by having them. We can have them without introspecting them. But when we introspect, our attentional access to them is direct in that it is unmediated by any experiential states. Experiences are in that sense self-presenting (Chisholm 1977). If this view is correct, then we are immune to a certain kind of error. When our perceptual experiences are illusory, when things are not as they appear, we can be misled into believing that they are as they appear. If, however, our conscious states are self-presenting, then there is no appearance/reality distinction that pertains to them. We thus cannot be misled about them by their appearing to us some way that they are not.

Some introspectionists maintain that an act of introspective awareness of a conscious state is direct in yet another sense: it is unmediated by any causal mechanism. If, however, introspective awareness of a conscious state involves believing something of the state (for example, that it is a pain), the question arises as to whether this of-ness connection requires causation. It seems like mystery-mongering to maintain that it is a primitive, fundamental relation. One view is that the relation is part-whole rather than causal: The conscious state is a constituent of the introspective belief. But there are constituents of the belief that the belief bears no of-ness relation to, for example, the concepts involved in it. So, the constituency must be of a special sort. Proponents of this view are under an obligation to explicate it. There is also the issue of whether such an account can allow for mistaken intro-



spective beliefs. These remain topics of investigation (see Chalmers 2003).

The more common view is that an introspective belief and the state introspected are linked by a causal mechanism. Causes and effects, however, must be “distinct existences,” and so capable of independent existence (Armstrong 1963). This causal view thus seems to entail that there could be a being with beliefs that it is in conscious states of various sorts on various occasions yet is never in such states. But perhaps there could be a silicon-based robot that is such a being—possessed of the relevant concepts but entirely devoid of sentience (McLaughlin and Tye 2003b). The shock of such a possibility is somewhat lessened if primary possession of concepts of conscious states requires acquaintance with such states (Peacocke 1998), so that the robot could possess them only in a secondary way—by communicative interaction with conscious beings that possess them in a primary way.

Another “independent existence” concern with the causal view is that it entails the possibility of beings who are in conscious states but lack the capacity to be introspectively aware of them, and so who are “self-blind” with respect to them (Shoemaker 1984b, 1984c). Introspectionists, however, maintain that introspective awareness of a conscious state consists of a belief that one is in the state, a belief formed by direct acquaintance with the state. Animals seem self-blind in the sense in question: they do not form beliefs about what conscious states they are in, for they lack the requisite concepts to do so. Indeed, animals do not introspect their conscious states; they are conscious, but not self-conscious. So, this sort of self-blindness may seem not to count against introspectionism. Nevertheless, there is a sense in which animals are aware of their pains or itches, for instance; that is why the dog yelps or scratches. Indeed it seems that their attention might be riveted on their sensation. It remains an open question whether the relevant mode of attention can be captured by a model of introspective attention as belief-formation or whether further distinctions are called for.

It has been claimed that when we try to direct our attention to our visual experience in order to introspect it, we seem to find ourselves only inspecting the scene before our eyes (Moore 1903; Harman 1990, Dretske 1994, 1999; Sturgeon 2000; Tye 2000). It is thus claimed that visual experience is phenomenologically “transparent” or “diaphanous.” And some philosophers claim that all conscious states are diaphanous (Tye 2000). The phenomenological thesis of transparency seems most plausi-

ble for visual experiences and least plausible for bodily sensations. But it is maintained that even when we attend to a toothache, our attention seems focused on a feature of the tooth itself, however alarming we may find that feature.

In the light of these phenomenological considerations, a “displaced-perception model” of first-person knowledge of experience has been proposed (Dretske 1994, 1999; Tye 2000). The leading idea in the visual case is this: when we are attentively aware that we are having a visual experience, our “awareness-that” is not based on direct awareness of the experience but rather on awareness of the scene before our eyes. Our awareness of the experience is indirect, because we are aware of it by being aware of the scene. Nevertheless, if we have mastered the concept of visual experience, we can come to be aware that we are having a certain visual experience, without recourse to consciously drawing inferences.

Hallucination seems to pose no problem for the phenomenological transparency thesis itself: Perhaps, whenever we visually hallucinate, we seem to be aware only of a scene. But hallucination poses a problem for the displaced-perception model if, when we (completely) hallucinate, we are not actually aware of any scene at all. If there are sense data (Ayer 1940), then we will actually be aware of a scene, even when we completely hallucinate, for sense data would constitute a scene. But the leading proponents of the displaced-perception model are physicalists and so deny that there are sense data. Proponents of the model have tried to accommodate hallucination by maintaining that in such a case one is aware of a type of scene, despite not being aware of any actual instance of it. Whether this model applies to visual experience and all conscious states remains a topic of controversy.

Our ordinary epistemic practices seem to rely not only on the presumption that our (sincere) first-person ascriptions of conscious states (for example, ‘I am in pain’) are *prima facie* true but also on the presumption that our first-person ascriptions of beliefs (for example, “I believe that P”), desires, and intentions are *prima facie* true. It has been claimed that the social-psychological data about confabulation shows the latter presumption to be unfounded. But, arguably, the data seem to show only that we have a tendency to confabulate when under pressure to explain how we arrived at our propositional attitudes or made choices; thus the data seems not to raise an unanswerable challenge to first-person authority. In any case, many contemporary philosophers claim that whatever role introspection may play in explaining our first-person authority as self-ascribers of conscious states, it

has little to do with our first-person authority concerning our propositional attitudes (Davidson 1984, 1986; Gallois 1996; Moran 2001).

Even if we indeed introspect conscious states (as these were characterized earlier), we do not introspect our beliefs, desires, or intentions. Indeed, we do not even introspect our attitudinal emotions (fear that *P*, anger that *Q*, relief that *R*, and so on). Such states can count as conscious, but only in the sense that they can have characteristic manifestations in consciousness; and (at best) we introspect only conscious states that manifest them. Thus, we may introspect an impulse, but not a desire; a feeling of anger, but not an attitude of anger; a thought that *P*, but not a belief that *P*. Indeed, to be aware of one's belief that *P* is just to be aware that one believes that *P*; and similarly for the other cases (Shoemaker 1994b). Just as we can typically know what we believe without observing our behavior, we can typically know what we believe without introspecting.

Moreover, although we sometimes know that we believe something as a result of assessing evidence that we do, such a case seems atypical. When we ask ourselves whether we believe that *P*, want *X*, or intend to *A*, we usually do not reflect on evidence concerning whether we believe that *P*, want *X*, or intend to *A*. Of course, we sometimes do that. But in response to the questions we typically reflect, respectively, on whether *P*, whether *X* has some attractive feature, and whether we ought to do *A* (Evans 1981, Gallois 1996, Moran 2001). Although we typically do that, reasons for believing that *P* is true are not reasons for believing that one believes that *P*; and reasons for believing that one ought to *A* are not reasons for believing that one intends to *A* (similarly for the desire case). Rather, they are, respectively, reasons to believe that *P* and reasons to intend to do *A*. So, the question of how such reflection leads to knowledge of our beliefs, desires, and intentions persists.

Philosophers who seek a role for introspection here will claim that, when we engage in such deliberative reflective reasoning, we can be introspectively aware of our occurrent thoughts. Philosophers who reject any role for introspection here will claim that even if we can indeed introspectively observe manifestations of propositional attitudes in consciousness and so have more "observational data" than others who can only observe manifestations of our attitudes in our overt verbal and nonverbal behavior, the fact that we have such additional observational data will not explain our first-person authority about our attitudes. Moreover, occurrently thinking that *P* is a mental act—indeed a basic mental

act: something we do, but not by doing something else (Moran 2001). Our knowledge of what we are occurrently thinking is knowledge of something that we are doing. Our distinctively characteristic knowledge of our basic actions may not be introspective. What explains first-person authority about our propositional attitudes and basic actions remains an open issue.

Many philosophers have related first-person authority about attitudes and actions to the fact that attitudes and actions (unlike bodily sensations, imagery, or sense experiences) can be rational or irrational. One view is that our practice of attributing propositional attitudes is essentially an interpretive practice governed (in part) by constitutive principles of rationality, and the presumption of first-person authority is required for interpretation to be possible (Davidson 1984, 1986). Another view is that the functional organization required to be a rational agent guarantees that a rational agent will, for the most part, be reliable in his or her beliefs about what propositional attitudes and experiences he or she has (Shoemaker 1994a, 1994b, 1994c). Yet another view seeks to explain our first-person authority in terms of rational commitment and first-person deliberation (Moran 2001). There are other very influential views (Burge 1988, 1993).

Belief, desire, intention, and occurrent thought are modes of intentionality; states of these (and other intentional) types have representational content. One issue is how one knows which of these (or other intentional) types a given intentional state falls under; another issue is how one knows what the content of the state is. Thus, there is, for instance, the issue of how one knows that one's belief that *P* is a belief (rather, than, say, a desire); and there is the issue of how one knows that one's belief is a belief that *P* (rather than a belief that something else is the case).

The leading contemporary theories of mental content are externalist theories, according to which the content of a mental state fails to supervene on intrinsic states of the subject (Putnam 1975, Burge 1979). On these views, two intrinsic duplicates (for example, an inhabitant of Earth and her doppelgänger on Twin Earth) could be in mental states with different contents. Some externalist theories hold that content depends on historical context (Dretske 1988), and according to others, it depends on social context (Burge 1979). There has been extensive debate about whether content externalism is compatible with our having first-person authority or privileged first-person knowledge concerning what we think. Some philosophers argue for incompatibilism (for

example, Boghossian 1989, 1997; McKinsey 1991). Some argue for compatibilism (for example, Davidson 1984, 1986; Burge 1988, 1993; Brucekner 1992; Heil 1992; Gallois 1996; Peacock 1998; Davies 1998; Gibbons 1996; Falvey and Owens 1994; McLaughlin and Tye 1998a, 1998b; McLaughlin 2000, 2003a; Brown 2004).

Here is an example of one of the leading incompatibilist lines of argument (McKinsey 1991, Boghossian 1997). For any of the content-externalists theories in question, there will be some contingent environmental proposition E such that E can be known only on the basis of empirical evidence, yet the theory will entail that it is a conceptual truth that if we are thinking that P, then E. Thus, if we could have privileged first-person knowledge that we are thinking that P, it follows that we would be able to infer that E and thereby come to know it on some basis other than empirical evidence. Some compatibilists have responded that the relevant contingent environmental propositions will be ones that can thereby be known on a basis other than empirical evidence, however surprising that might be (Warfield 1998, Sawyer 1998). But by far the more prevalent compatibilist response is to try to show that combinations of the relevant content-externalist and privileged self-knowledge theses do not lead to this result (Brueckner 1992; Davies 1998; McLaughlin and Tye 1998a, 1998b; 2003a).

**See also** Behaviorism; Consciousness; Descartes, René; Freud, Sigmund; Introspection; Intuition; Memory; Perception; Personal Identity; Ryle, Gilbert; Self; Socrates; Unconscious.

### Bibliography

- Alston, W. P. "Varieties of Privileged Access." *American Philosophical Quarterly* 8 (1971): 223–241.
- Armstrong, D.M. "Is Introspective Knowledge Incorrigible?" *Philosophical Review* 62 (1963): 417–32.
- Ayer, A. J. *The Foundations of Empirical Knowledge*. London: Macmillan, 1940.
- Boghossian, P. "Content and Self-Knowledge." *Philosophical Topics* 1 (1989): 5–26.
- Boghossian, P. "What the Externalist Can Know A Priori." *Proceedings of the Aristotelian Society* 97 (1997): 161–175.
- Broad, C. D. *The Mind and Its Place in Nature*. London: Routledge and Kegan Paul, 1925.
- Brown, J. *Anti-Individualism and Knowledge*. Cambridge, MA: MIT Press, 2004.
- Brueckner, A. "What an Anti-Individualist Knows A Priori." *Analysis* 52 (1992): 111–118.
- Burge, T. "Content Preservation." *The Philosophical Review* 102 (1993): 457–488.
- Burge, T. "Individualism and Self Knowledge." *The Journal of Philosophy* 83 (1988): 649–663.
- Burge, T. "Individualism and the Mental." *Midwest Studies in Philosophy* 4 (1979): 73–121.
- Chalmers, D. "The Content and Epistemology of Phenomenal Belief." In *Consciousness: New Philosophical Essays*, edited by Q. Smith and A. Jokic. Oxford, 2003.
- Chisholm, R. *Theory of Knowledge*. 2nd ed. Prentice Hall, 1977.
- Davidson, D. "First Person Authority." *Dialectica* 38 (1984): 101–111.
- Davidson, D. "Knowing One's Own Mind." *Proceedings of the American Philosophical Association* 60 (1986): 441–458.
- Davies, M. "Externalism, Architecturalism, and Epistemic Warrant." In *Knowing Our Own Minds*, edited by C. Wright, B. Smith, and C. Macdonald. Oxford, 1998.
- Dennett, D. *Consciousness Explained*. Boston: Little, Brown & Company, 1991.
- Descartes, R. *The Philosophical Writings of Descartes*. Vols. 1 and 2. Translated by J. Lottingham, R. Stoothoff, and D. Murdoch. Cambridge, U.K.: Cambridge University Press, 1985.
- Dretske, F. "Introspection." *Proceedings of the Aristotelian Society* 94 (1994): 263–278.
- Dretske, F. "The Mind's Awareness of Itself." *Philosophical Studies* 95 (1999): 103–124.
- Evans, G. *The Varieties of Reference*. New York: Oxford University Press, 1982.
- Falvey, K. and Owens, J. "Externalism, Self-Knowledge, and Skepticism." *The Philosophical Review* 103 (1994): 107–137.
- Freud, S. *The Interpretation of Dreams*. 3rd ed. Translated by A. A. Brill. London: Allen and Unwin, 1923.
- Gallois, A. *The World Without, the Mind Within: An Essay on First-Person Authority*. Cambridge, U.K.: Cambridge University Press, 1996.
- Gertler, B. "Self-Knowledge." *The Stanford Encyclopedia of Philosophy*, Spring 2003 edition, edited by E. N. Zalta. Available from <http://plato.stanford.edu/archives/spr2003/entries/self-knowledge/>.
- Gibbons, J. "Externalism and Knowledge of Content." *The Philosophical Review* 105 (1996): 287–310.
- Goldman, A. *Pathways to Knowledge: Private and Public*. New York: Oxford University Press, 2002.
- Harman, G. "The Intrinsic Quality of Experience." In *Philosophical Perspectives*, edited by J. Tomberlin. Atascadero, CA: Ridgeview, 1990.
- Heil, J. *The Nature of True Minds*. Cambridge, U.K.: Cambridge University Press, 1992.
- Hill, C. S. *Sensations*. New York: Cambridge University Press, 1991.
- Locke, J. *An Essay concerning Human Understanding*. Oxford, 1690/1975.
- Ludlow, P., and Martin, N. eds. *Externalism and Self-Knowledge*. Stanford, CA: CSLI Publications, 1998.
- Lycan, W. G. *Consciousness and Experience*. Cambridge, MA: MIT Press, 1996.
- Macdonald, C. "Externalism and Authoritative Self-Knowledge." In *Knowing Our Own Minds*, edited by C. Wright, B. Smith, and C. Macdonald. Oxford: Oxford University Press, 1998.

- Macdonald, C. "Shoemaker on Self-Knowledge and Inner Sense." *Philosophy and Phenomenological Research* 59 (1999): 711–738.
- McKinsey, M. "Anti-Individualism and Privileged Access." *Analysis* 51 (1991): 99–121.
- McLaughlin, B. P., and A. Rorty, eds. *Perspectives on Self-Deception*. Berkeley: University of California Press, 1988.
- McLaughlin, B. P., and M. Tye. "Color, Consciousness, and Color Consciousness." In *Consciousness: New Philosophical Essays*, edited by Q. Smith and A. Jokic. Oxford: Clarendon Press, 2003.
- McLaughlin, B. P., and M. Tye. "Externalism, Twin Earth, and Self-Knowledge." In *Knowing Our Own Minds*, edited by C. Wright, B. Smith, and C. Macdonald. Oxford: Clarendon Press, 1998a.
- McLaughlin, B. P., and M. Tye. "Introspecting Thoughts." *Facta Philosophica* 3 (2001): 77–83.
- McLaughlin, B. P., and M. Tye. "Is Content Externalism Compatible with Privileged Access?" *The Philosophical Review* (1998b): 349–380.
- McLaughlin, B. P., and M. Tye. "McKinsey's Challenge, Warrant Transmission, and Skepticism." In *New Essays on Semantic Externalism and Self-Knowledge*. Edited by S. Nuccetell, 79–96. Cambridge, MA: MIT Press, 2003a.
- McLaughlin, B. P., and M. Tye. "A Naturalist Phenomenal Realist Response to Block's Harder Problem." *Philosophical Issues* 13 (2003b): 163–204.
- McLaughlin, B. P., and M. Tye. "Skepticism, Externalism, and Self-Knowledge." *The Aristotelian Society*, supp. vol. 74 (2000): 93–118.
- Moore, G. E. "The Refutation of Idealism." In *Philosophical Papers*. Routledge and Kegan Paul, 1903.
- Moran, R. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton, NJ: Princeton University Press, 2001.
- Parke, H. W. *A History of the Delphic Oracle*. Oxford: B. Blackwell, 1939.
- Peacocke, C. "Conscious Attitudes, Attention, and Self-Knowledge." In *Knowing Our Own Minds*, edited by C. Wright, B. Smith, and C. Macdonald. Oxford: Clarendon Press, 1998.
- Putnam, H. "The Meaning of 'Meaning.'" In *Mind, Language and Reality*. New York: Cambridge University Press, 1975, 215–271.
- Ryle, G. *The Concept of Mind*. New York: Hutchinson's University Library, 1949.
- Sawyer, S. "Privileged Access to the World." *Australasian Journal of Philosophy* 76 (1998): 523–533.
- Sellars, W. "Empiricism and the Philosophy of Mind." In *Perception, and Reality*, edited by W. Sellars. London: Routledge and Kegan Paul, 1963.
- Shoemaker, S. *Identity, Cause, and Mind: Philosophical Essays*. Cambridge, U.K.: Cambridge University Press, 1994a.
- Shoemaker, S. "Introspection." in S. Guttenplan ed. *A Companion to the Philosophy of Mind*. Cambridge, 1994c, 395–400.
- Shoemaker, S. "Self-Knowledge and 'Inner Sense.'" *Philosophy and Phenomenological Research* 54 (1994b): 249–314.
- Sturgeon, S. *Matters of Mind: Consciousness, Reason, and Nature*. New York: Routledge, 2000.

Tye, M. *Consciousness, Color, and Content*. Cambridge, MA: MIT Press, 2000.

Warfield, T. "A Priori Knowledge of the World: Knowing the World by Knowing Our Minds." *Philosophical Studies* 92 (1998): 127–147.

Williamson, T. *Knowledge and Its Limits*. Oxford: Oxford University Press, 2000.

Wright, C. "Self-Knowledge: The Wittgensteinian Legacy." In *Knowing Our Own Minds*, edited by C. Wright, B. Smith, and C. Macdonald. Oxford: Clarendon Press, 1998.

Ziff, P. "About Behaviorism." *Analysis* XVIII (1958).

**Brian P. McLaughlin (2005)**

## SELF-PREDICTION

In recent years philosophers have produced arguments designed to prove that not all human behavior can be predicted or otherwise known in advance, and these arguments have been taken to be relevant to the problem of freedom of the will as well as to the question whether there can be genuine behavioral sciences. Specifically, it is argued that in certain circumstances it is logically impossible that one should come to know decisions, and actions for whose occurrence decisions are necessary conditions, in advance of the occurrence of such decisions. This has been interpreted as a refutation of determinism.

Two antipredictive arguments will be presented separately, and later their import when taken together will be discussed. The first concerns the scientific defectiveness of predictions that influence the predicted event, and the second concerns the logical impossibility of a person's knowing now what he will decide only at some future time.

## INFLUENCE OF PREDICTIONS

It is a familiar fact that some prophecies and predictions are self-fulfilling in the sense that the prediction itself produces the predicted event—for example, when all the stock market tip sheets predict that stock *x* will drop sharply in the next few weeks. We also know, for similar reasons, that some predictions are self-defeating. For example, Jones predicts that he will, as usual, take the easy way out of a difficulty, but then, to prove to himself that he can do better, he does just the opposite. This prediction affected his deliberation and caused him to make a decision opposite to the one he had predicted. Now, the argument that follows does not maintain that a person's predictions of his own future decisions are necessarily or always self-defeating; instead, it maintains that it is logically impossible that by considering causes a person

should come to know that his final prediction of what he will decide is not self-defeating, and it maintains that the attempt to achieve such knowledge involves an infinite regress. In other words, this antipredictive argument purports to prove that predictions of one's own future decisions on the basis of antecedent causal conditions cannot possibly be scientifically complete.

It is necessary to state some assumptions and restrictions required by the argument. The first assumption is that decisions are events and hence are the sorts of things that can be caused; many philosophers would reject this assumption. Second, the argument concerns only causal knowledge of future decisions, by which is meant predictions derived with scientific adequacy from what one knows to be all the relevant antecedent causes of the decision, as distinct from predictions not known to be based on all the relevant causes and which consequently yield only a likelihood of the decision's occurrence. Finally, the argument aims to prove only that it is logically impossible for a person to have causal knowledge of his own decision in advance of making such a decision.

Let us assume, then, that some set of circumstances *C* is causally sufficient for a person *S* to make decision *D* and that *S* has unlimited knowledge of past circumstances and relevant causal laws. Can *S* come to know that *C* is sufficient for *D*? *S* may come to make a prediction *P* that past circumstances *C* are sufficient for *D*. We have supposed that as a matter of fact *C* is causally sufficient for *D*, but *S* nevertheless cannot know that this is so unless he also knows that there are no contrary causes. That is, before *S* can know that *C* is sufficient for *D* he must also know that there is no other circumstance which, together with *C*, is sufficient for not-*D*. One such probable cause of not-*D* is the prediction itself. Therefore, *S* cannot know that *C* is sufficient for *D* unless he knows that it is false that

(1) *C* plus *P* are causally sufficient for not-*D*.

*S* has been allowed unlimited knowledge of past circumstances and relevant causal laws, hence *S* can know that (1) is false, that is, he can know that making the prediction will not cause him to make a different decision. It does not follow, however, that *S* now can know that *C* is sufficient for *D*, for the same problem recurs: *S*'s knowledge that (1) is false, which we will call *P*<sub>1</sub>, is a new datum and is itself a possible cause of not-*D*. Therefore, *S* cannot know that *C* is sufficient for *D* unless he knows that it is false that

(2) *C* plus *P*<sub>1</sub> are causally sufficient for not-*D*.

And *S*'s knowledge that (2) is false, or this knowledge plus his feelings or attitudes toward (2), constitute a further possible contrary cause, *P*<sub>2</sub>. Thus, an infinite regress arises, within which the agent's prediction on the basis of some evidence *C* or his revision of the prediction or his final thoughts about the prediction are relevant data in addition to the data upon which the prediction was based. *S*'s calculating of causes cannot possibly "catch up" with the number of possible causes that must be examined if the prediction is to be scientifically complete, for the final results obtained cannot themselves also be part of the basis of one's prediction.

When one attempts to predict a supernova, it is true that in this case, too, the final prediction arrived at is necessarily excluded from the data upon which the prediction is based. However, although it is logically possible that predictions or thoughts about predictions can produce or impede a supernova, it is not scientifically possible that they do so. Therefore, the infinite regress argument is no obstacle to knowledge of, for example, scientific laws or stellar events but concerns only particular events that can be produced or prevented by human agency. And it is clearly applicable to attempted predictions of one's own decisions because we know that speculations and predictions about what one is likely to decide are always among the conditions most likely to be determinative of what one will in fact decide.

#### COUNTERARGUMENTS FAVORING DETERMINISM.

The view that this first antipredictive argument casts doubt on determinism may be challenged in a number of ways:

(a) The argument presents no obstacle to the existence of a complete causal explanation of one's own past decisions.

(b) There is no logical obstacle to a person's predicting a future decision of someone other than himself, although such prediction does confront a methodological difficulty. That is, suppose that *A* predicts a future decision of *B*'s and resolves not to tell *B* the prediction. Then it appears that *A* must also predict something about himself; namely, that he will not later decide to revoke his past decision and tell *B*, after all—and this, according to the infinite regress argument, *A* cannot possibly do. One complication here is the question whether the regress argument precludes *A*'s predicting that he will make no decisions at all during a certain future period; if the regress argument does not preclude this, then *A* can predict that he will not change his mind and tell the original prediction to *B*. But in any case the solution seems to lie in

having *A* make his prediction of *B*'s decision from a dungeon or a distant planet or in such a way that he has no time to communicate with *B* in advance of *B*'s making his decision; that is, perhaps it is sufficient that it be physically (although not logically) impossible that *A* should ruin the impeccable scientific basis of his prediction by telling *B*.

(c) The regress argument shows no peculiarity of human or even of sentient beings. For it is easy to imagine a simple machine, for which no one would dream of claiming free will or moral responsibility, the behavior of which could not possibly be predicted in circumstances similar to those previously described. We need only suppose that the machine can do two things, *x* and *y*, that a prediction of either of these things, punched into a card, can be inserted in the machine, and that we announce our predictions of what the machine will do by inserting appropriately punched cards into the machine. The machine is built to do *x* when fed the prediction "machine will do *y*" and to do *y* when fed the prediction "machine will do *x*." The situation in which a prediction of a person's decision is defective is fully as artificial as this, and in each situation the prediction is defective for the same reason. In each case, given the causal hypothesis, one can in principle make a scientifically impeccable prediction of what will occur only if neither the person nor the machine is allowed to be influenced by the prediction. Meaning "*y*" when one inserts the card saying "machine will do *x*" into the machine is equivalent to telling a person he will decide not-*D* when one knows that telling him this will cause him to decide *D*.

It can be argued that the first antipredictive argument shows only that given the causal hypothesis, it is still possible to make predictions competently and incompetently and that one of countless ways in which one can make predictions incompetently is to allow one's prediction to disturb the system that one is trying to predict. However, although it may be the case that the self-defeating prophecy and the self-fulfilling prophecy are equally explicable and, in general, equally avoidable phenomena, it appears that the special situation in which the self-defeating prophecy is unavoidable is important to us—namely, the situation in which we attempt to predict our own decisions. The regress argument also poses a methodological problem for social scientists who wish to circulate predictions of human behavior, but it does not show that there is any event that in principle cannot be predicted.

## LOGICAL IMPOSSIBILITY OF SELF-PREDICTION

The second antipredictive argument appears to follow from the analytic truth that one cannot know now what, by hypothesis, one will not know until some later time. Thus, one form of this argument (see Karl Popper, "Post-script: After Twenty Years") maintains that exact historical prophecy is incompatible with the fact of advancing knowledge. That is, it is impossible to predict the future decisions and actions of people because these future decisions and actions will be formed and done on the basis of knowledge that, by hypothesis, no one now possesses.

Another form of the argument maintains that it is logically impossible for a person to know what he will decide to do before he actually makes his decision (see Stuart Hampshire, *Thought and Action*; Carl Ginet, "Can the Will Be Caused?"; and D. F. Pears, *Freedom and the Will*). It is claimed that if a person knows or thinks he knows what he will try to do tomorrow, then either he has already decided what he will try to do or he believes that what he will try to do is not up to him. In neither of these two cases can he decide what he will try to do, for in each case there is nothing for him to decide. Decision is making up one's mind about what one will try to do or about what one will acquiesce in; therefore, to say that one will decide tomorrow appears to entail that there is something one will know then and which, by hypothesis, one does not know now.

However, there is a difficulty here. What is it that one knows as a result of decision and that one cannot know prior to the decision? From the fact that a person has decided to do something, it does not follow that he knows what he will do or try to do in the future. Decision does not give one knowledge of anything that will occur in the future because the mere fact that a person has decided does not ensure that he will not falter, change his mind, or die tomorrow. Hence, it appears to be mistaken to assume that because decision entails ignorance prior to decision, this ignorance is of something which one will know later as a result of decision; what one comes to know when one decides is nothing in addition to the decision itself and not any fact about the future. The reason for this appears to be that "decision" is an intentional concept.

Sometimes a person claims to know what a future decision of his will be, and various explanations of his supposed mistake can be made: (a) He has already decided, and he confuses with the act of decision itself some future reaffirmation, announcement, or implementation of his decision. (b) He has tentatively decided and plans at the last moment to reappraise his decision, but he

thinks that he knows the result of that reappraisal because of his tentative decision. In this case, if he does not deliberate again at the last moment, then he merely reaffirms what he has already decided, and if he does deliberate again, then it is impossible that he should know in advance the result of his deliberation, even though this new decision agrees with his earlier tentative decision. (c) He construes a future reaffirmation of a decision already made to be a new decision because its time, place, or context differs from that in which he first decided. (d) He confuses a guess, likelihood, or probability with knowledge of his future decision.

It has also been claimed (for example, by Richard Taylor, in “Deliberation and Foreknowledge”) that if a person knows or thinks he knows what he will do in the future, then it is impossible for him to deliberate about what he will do, for deliberation also presupposes ignorance. “Jones is deliberating whether to do *x*” entails “Jones does not know whether or not he will do *x*.” But here a distinction must be made between the agent’s belief or knowledge that he will do a particular act in the future and the agent’s belief or knowledge that this particular act he will do is in some sense not up to him. If a person believes that he will do *x*, he cannot deliberate whether to do *x*, even though he believes that he will do *x* freely, that what he does is up to him. On the other hand, if a person believes that what he will do is not up to him, then he cannot deliberate whether to do *x*, even though he lacks knowledge or belief about what he will do. Hence, although it has been claimed that both foreknowledge and lack of freedom preclude deliberation and decision, these claims nevertheless require separate argument, and only foreknowledge is relevant to self-prediction and the paradoxes thereof.

It might be thought that the two antipredictive arguments are not truly distinct, and indeed some philosophers have written as though these arguments were but two approaches to the same logical point. But they are distinct, except insofar as they can be put to similar purposes. The first argument applies to all predictions that can causally influence the events predicted, whether these events happen to be decisions, revolutions, or stock market trends. It is thus broader in scope and does not require that the event also be of that special sort which, in certain circumstances, is logically impossible to know in advance. The second argument attacks the very idea of foreknowledge, however obtained, of occurrences that entail prior ignorance and does not, as does the first argument, attack the scientific adequacy of predictions that can influence the predicted events.

## LOGICAL IMPOSSIBILITY OF CAUSING DECISIONS

Many philosophers would maintain that if some set of antecedent conditions is causally sufficient for the occurrence of an event, then it is logically possible that the event be predicted or known prior to its occurrence. From this claim, together with the second antipredictive argument, can be constructed the following argument that attempts to prove that it is logically impossible that decisions have causes (see Ginet, *op. cit.*): If it is logically possible for a decision to be caused, then it is logically possible for a person to know what his own decision will be before he makes his decision; it is not logically possible for a person to know what his own decision will be before he makes his decision; therefore, it is not logically possible for a decision to be caused.

This argument is, in the following way, of more apparent relevance to the traditional problem of freedom of the will and in particular to a theory of human agency: Let us suppose that decisions are necessary conditions for the occurrence of certain actions, and let us suppose further that decisions are part of the causes of such actions. If so, then any set of causes sufficient for the occurrence of such an action must include a decision as part of the set, for whatever is sufficient for something to occur must include everything necessary for that thing to occur. But the decision, by the preceding argument, is uncaused, and therefore no set of causes existing prior in time to the decision can be sufficient for the occurrence of the action. The decision can thus be viewed as a partial, uncaused cause of the action, which, together with ordinary causes, is sufficient for the occurrence of the action.

Difficulties of the following sort have been raised against the argument that maintains that it is impossible that decisions be caused: First, it has been doubted that it follows from the causal hypothesis that it is possible for a person to predict his own decisions; for the possibility of predictability in principle need not include the possibility of predictability in all possible circumstances (see A. J. Stenner, “On Predicting Our Future”). As we have seen, it is not obvious that paradoxes arise when we suppose someone to predict decisions of persons other than himself. Second, a premise of this argument maintains that from the hypothesis that decisions are caused, it follows that one could in principle make a scientifically adequate prediction, based on knowledge of antecedent causes, of one’s own future decision. But the first of the two antipredictive arguments claims that this does not follow at all, because it is impossible to establish that one’s prediction has no contrary influence on the predicted event.

That is, the first antipredictive argument, if sound, shows that the causal hypothesis does not entail the apparent absurdity that in principle one could, by considering antecedent conditions and relevant causal laws, come to know one's own decisions in advance.

### **Bibliography**

- Canfield, J. "Knowing about Future Decisions." *Analysis* 22 (6) (1962): 127–129.
- Fain, Haskell. "Prediction and Constraint." *Mind* 67 (1958): 366–378.
- Ginet, Carl. "Can the Will Be Caused?" *Philosophical Review* 71 (1) (1962): 49–55.
- Hampshire, Stuart. *Thought and Action*. London: Chatto and Windus, 1960.
- Hampshire, Stuart, and H. L. A. Hart. "Decision, Intention and Certainty." *Mind* 67 (265) (1958): 1–12.
- Herbst, P. "Freedom and Prediction." *Mind* 66 (1957).
- Lucas, J. R. "Freedom and Prediction." Part 1. *PAS*, Supp. 41 (1967): 163–172.
- Lyon, A. "The Prediction Paradox." *Mind* 68 (1959): 510–517.
- MacKay, D. M. "Brain and Will" (BBC Third Program, 1957). Printed in *A Modern Introduction to Philosophy*, edited by Paul Edwards and Arthur Pap. New York: Free Press, 1965.
- MacKay, D. M. "On the Logical Indeterminacy of a Free Choice." *Mind* 69 (1960).
- Moran, Richard. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton, NJ: Princeton University Press, 2001.
- Moran, Richard. "Self-Knowledge: Discovery, Resolution, and Undoing." *European Journal of Philosophy* 5 (1997): 141–161.
- Murdoch, Iris, Stuart Hampshire, Patrick Gardiner, and D. F. Pears. "Freedom and Knowledge." In *Freedom and the Will*, edited by D. F. Pears, 80–104. London, 1963.
- Perry, David. "Prediction, Explanation, and Freedom." *Monist* 49 (1965): 234–247.
- Popper, Karl. "Postscript: After Twenty Years." In his *Logic of Scientific Discovery*. London: Hutchinson, and New York: Basic, 1959.
- Rabinowicz, Wlodek. "Does Practical Deliberation Crowd Out Self-Prediction?" *Erkenntnis* 57 (2002): 91–122.
- Schlesinger, George. "The Unpredictability of Free Choices." *British Journal for the Philosophy of Science* 25 (1974): 209–221.
- Stenner, A. J. "On Predicting Our Future." *Journal of Philosophy* 61 (14) (1964): 415–428.
- Stocker, Michael. "How to Prevent Self-Prediction." *Journal of Philosophy* 65 (1968): 475–476.
- Taylor, Richard. "Deliberation and Foreknowledge." *American Philosophical Quarterly* 1 (1) (1964).

**Andrew Oldenquist (1967)**

*Bibliography updated by Benjamin Fiedor (2005)*

## SELLARS, ROY WOOD

(1880–1973)

Roy Wood Sellars, the American critical realist, taught philosophy at the University of Michigan. Although he was never as well known outside philosophical circles as some of his contemporaries, after the publication of his first book, *Critical Realism*, in 1916, Sellars maintained a substantial reputation among his fellow philosophers as a vigorously independent thinker. His thought was rigorous and critical; he never yielded to the fashionable movements of the day but steadfastly pursued his own original insights into basic philosophical problems.

The core of Sellars's philosophy is epistemological. He is concerned with showing that the critical realism of the philosopher is related to the "natural realism" of the "plain man." The philosopher reflects on the plain man's uncritical view of knowledge, which he clarifies and refines so that it is philosophically justifiable, but he does not vitiate its essential insistence upon the independence of the object of knowledge. The most significant element in Sellars's vindication of realism is his revision of the theory of perception, which he describes as a process of interpretation of *sensa*, as mediated by factors both external and internal to the perceiving subject. This view of perception avoids both the simplistic claim of natural realism that things reveal themselves directly in perception and the subjectivist claim that the objects of perception are ideas rather than things. Knowledge, too, is a complex process and occurs at various levels of complication. Its ultimate biological source is to be found in the adjustment of the organism to its environment; its ultimate outreach is in scientific knowledge, which replaces the relativity of individual perspectives by close approximations to exact measurement. Whether on the implicit organic level or on the highly explicit and self-critical scientific level, we know that we know when the content of our beliefs corresponds to the externally observed state of affairs.

Working from this epistemological position, Sellars developed an evolutionary cosmology and a materialistic ontology, carrying on his insight that there are levels, or "gradients," of being. Even the higher levels like life and mind, which emerge under most favorable conditions, are, however, physical systems. Sellars's materialism is nonreductive, but he insists that "life is not a nonnatural force coming from outside, but a term for the new capacities of which nature has found itself capable." On the valuational side, Sellars argues from these positions to a humanistic theory of ethics and religion (he was one of



the major contributors to the composition of the Humanist Manifesto of 1933) and to a politics of democratic socialism.

**See also** Critical Realism.

## Bibliography

### WORKS BY SELLARS

- Critical Realism*. Chicago: Rand McNally, 1916.  
*The Next Step in Democracy*. New York: Macmillan, 1916.  
*The Next Step in Religion*. New York: Macmillan, 1918.  
*Evolutionary Naturalism*. Chicago: Open Court, 1922.  
*Religion Coming of Age*. New York: Macmillan, 1928.  
*The Philosophy of Physical Realism*. New York: Macmillan, 1932.  
*The Philosophy of Physical Realism*. New York: Russell & Russell, 1966.  
*Lending a Hand to Hylas*. N.p., 1968.  
*Reflections on American Philosophy from Within*. Notre Dame, IN: University of Notre Dame Press, 1969.  
*The Principles, Perspectives, and Problems of Philosophy: An Exploration in Depth*. New York: Pageant Press International, 1970.  
*Patterns and Political Horizons*. Nashville, TN: Aurora, 1970.  
*Social Philosophy and the American Scene: Materialism and Human Knowing*. New York: Oriole Editions, 1970.  
*Principles of Emergent Realism: Philosophical Essays*. St. Louis, MO: W.H. Green, 1970.  
*Neglected Alternatives: Critical Essays*. Lewisburg: Bucknell University Press, 1973.

### WORKS ON SELLARS

- Bahm, Archie, R. M. Chisholm et al. "A Symposium in Honor of Roy Wood Sellars." *Philosophy and Phenomenological Research* 15 (1954): 1–97.  
 Reck, Andrew. *Recent American Philosophy*. New York: Pantheon, 1964.

**J. L. Blau (1967)**

*Bibliography updated by Michael J. Farmer (2005)*

## SELLARS, WILFRID

(1912–1989)

Wilfrid Stalker Sellars, an American philosopher and teacher, was born in Ann Arbor, Michigan, the son of Roy Wood Sellars, the American critical realist who taught at the University of Michigan. Wilfrid Sellars's early education took place in the United States and in France, where he attended the lycées Montaigne and Louis le Grand; it was continued at the University of Michigan (BA, 1933), the University of Buffalo (MA, 1934), and Oxford University, where he was a Rhodes scholar and received a BA with first-class honors in philosophy, politics, and eco-

nomics. He received an MA from Oxford in 1940. After a year at Harvard University he began his career as a teacher of philosophy in 1938 at the University of Iowa. During the war he spent several years as an officer in the Naval Reserve, and in 1946 he went to the University of Minnesota, where he eventually became professor of philosophy, chairman of the philosophy department, founding co-editor of the journal *Philosophical Studies*, and a member of Herbert Feigl's Minnesota Center for the Philosophy of Science. In 1959 he joined the faculty of Yale University, and in 1963 he moved to the University of Pittsburgh, where he became University Professor of Philosophy and Research Professor of the Philosophy of Science. Apart from numerous interludes as a visiting professor at other institutions, he remained at Pittsburgh until his death.

Although Sellars became an extremely prolific writer, in the early years of his career he had great difficulty putting his ideas on paper. His first scholarly essay, third in his list of publications, was "Realism and the New Way of Words"; it underwent seventeen major revisions, Sellars said in his "Autobiographical Reflections" (1975), before it finally appeared in print. In spite of its striking originality, his early work was strongly influenced by the logical empiricist movement, particularly by the work of Rudolf Carnap; in one essay, "Epistemology and the New Way of Words," he declared that philosophy "is properly conceived as the pure theory of empirically meaningful languages." From the vantage point of the early twenty-first century, perhaps the most significant of his early essays would be "Concepts as Involving Laws and Inconceivable without Them" (1948) and "A Semantical Solution of the Mind–Body Problem" (1953). Both show him to have been well ahead of his time in analytic philosophy. In the former he offered a clarification of necessity and natural law that anticipated the treatment of these notions in recent possible-world semantics, and in the latter he developed a distinctly functionalist view of intentional states. (The early essays discussed here are included in the volume *Pure Pragmatics and Possible Worlds: The Early Essays of Wilfrid Sellars*, edited by J. F. Sicha.)

Sellars's best-known philosophical work is the lengthy essay "Empiricism and the Philosophy of Mind," included in Sellars' *Science, Perception, and Reality*. This essay originated in lectures that Sellars gave in 1956 attacking what he called "the myth of the given." The cluster of ideas making up this doctrine was, he thought, the source of important errors in both the theory of knowledge and the philosophy of mind; by exposing the doc-

trine as a myth, he hoped to lay the groundwork for an acceptable form of empiricism and for a proper understanding of mental and sensory phenomena. The basic epistemic error prompted by the myth was the idea that empirical knowledge rests on a foundation of certain truth that is simply given to the mind—that is, knowable without inference—and provides the ultimate evidence for anything knowable by inference. The root error in the philosophy of mind prompted by the myth was the conviction that, merely by having sensory experiences and conscious thoughts, people gain theoretically satisfactory conceptions of those experiences and thoughts. These corresponding errors are related by the belief, commonly held by those who accept the myth, that foundational empirical knowledge concerns the sensory and psychological items, the mere having of which supposedly results in their being adequately conceived of or understood.

In attacking the errors he saw in the myth Sellars defended the view that empirical knowledge cannot have a foundation—that the supposedly basic knowledge of psychological fact presumed by the myth cannot exist independently of general knowledge relating psychological experience to linguistic and other behavior—and that theoretically adequate conceptions of anything can be obtained only by a process of learning and can be known to be adequate only by reference to scientific theorizing about the sensory and cognitive capabilities of human beings. He argued that “empirical knowledge ... is rational not because it has a *foundation* but because it is a self-correcting enterprise which can put *any* claim in jeopardy, though not *all* at once” (1991, pp. 127–196). As for commonsense sensory and psychological concepts, he argued that it is illuminating to think of them as resulting from an attempt to explain intelligent, nonhabitual human behavior by postulating appropriate “inner episodes” in substantially the way that theoretical scientists explain facts about observable objects by postulating unobservable microcauses. In arguing this point he added that, when concepts of such inner episodes are developed, people can learn to use them in making first-person reports of what they are experiencing. Seen this way, psychological concepts are fundamentally intersubjective rather than private, and they are as subject to revision as any concept of theoretical science.

In “Philosophy and the Scientific Image of Man” (1960), also included in *Science, Perception, and Reality*, Sellars developed the thesis that, although theoretical science is a natural development of commonsense thought about the world, it is not evidentially dependent upon it.

Like David Hume, Sellars thought that scientific thinking yields a theoretical picture of humans in the world that is incompatible with the commonsense—or, as he called it, the “manifest”—image of the same reality. These clashing images are not on a par, he thought; in purely descriptive respects, the scientific image is an improvement upon the manifest image, containing “successor concepts” to commonsense counterparts. (Water, on this view, is not identical with H<sub>2</sub>O; the technical concept of H<sub>2</sub>O applies to a common ingredient in most puddles, wells, clouds, and seas—one that is not accurately singled out by any commonsense concept.) A philosophically adequate picture of humans in the world is not fully descriptive, however; it is partly normative. Working out such a picture is an important philosophical task that has yet to be accomplished: the scientific image is not yet complete, and serious problems exist about how some normative matters can be incorporated into a significantly different image.

In later writings Sellars worked out highly original ideas on most central fields of philosophy. He produced, as Johanna Seibt (1990) observed, a unique scheme of “full scope nominalism,” which purports to demonstrate the expendability of abstract entities for all their supposed explanatory functions; he worked out (he was the first to do so) a sophisticated “conceptual role” semantics: he developed a neo-Kantian view of moral obligation and the moral point of view; and he had original things to say about central figures and issues in the history of philosophy. At a time when systematic philosophy was decidedly out of fashion, Sellars pursued the synoptic vision of humans in the world that Plato spoke of in the Republic. In parody of Kant he liked to tell his students that in philosophy analysis without synthesis must be blind.

**See also** Carnap, Rudolph; Empiricism; Functionalism; Philosophy of Mind.

## Bibliography

### WORKS BY SELLARS

- “Autobiographical Reflections.” In *Action, Knowledge, and Reality: Critical Studies in Honor of Wilfrid Sellars*, edited by H.-N. Castañeda. Indianapolis: Bobbs-Merrill, 1975.
- Philosophical Perspectives: History of Philosophy*. Atascadero, CA: Ridgeview, 1977.
- Philosophical Perspectives: Metaphysics and Epistemology*. Atascadero, CA: Ridgeview, 1977.
- Pure Pragmatics and Possible Worlds: The Early Essays of Wilfrid Sellars*, edited by J. F. Sicha. Atascadero, CA: Ridgeview, 1980.
- Science, Perception, and Reality*. Atascadero, CA: Ridgeview, 1991.
- Naturalism and Ontology*. Atascadero, CA: Ridgeview, 1992.

*Science and Metaphysics*. Atascadero, CA: Ridgeview, 1992.

#### WORKS ON SELLARS

Delaney, C. F., et al. *The Synoptic Vision: Essays on the Philosophy of Wilfrid Sellars*. South Bend, IN: Notre Dame, 1977.

DeVries, Willem A., and Timm Triplett. *Knowledge, Mind, and the Given: A Reading of Sellars' "Empiricism and the Philosophy of Mind."* Indianapolis: Hackett, 2000.

Pitt, Joseph C., ed. *The Philosophy of Wilfrid Sellars*. Boston: Reidel 1978.

Seibt, J. *Properties as Processes: A Synoptic Study of Sellars' Nominalism*. Atascadero, CA: Ridgeview, 1990.

**Bruce A. Aune (1996, 2005)**

## SEMANTICS

Semantics is the study of meaning. More specifically, semantics is concerned with the systematic assignment of meanings to the simple and complex expressions of a language. The best way to understand the field of semantics is to appreciate its development through the twentieth century. In what follows, that development is described. As will be seen, advances in semantics have been intimately tied to developments in logic and philosophical logic.

Though there were certainly important theories, or proto-theories, of the meanings of linguistic expressions prior to the seminal work of the mathematician and philosopher Gottlob Frege, in explaining what semantics is it is reasonable to begin with Frege's mature work. For Frege's work so altered the way language, meaning and logic are thought about that it is only a slight exaggeration to say that work prior to Frege has been rendered more or less irrelevant to how these things are currently understood.

In his pioneering work in logic *Begriffsschrift, eine der arithmetischen nachgebildete Formalsprache des reinen Denkens*, which was published in 1879, Frege literally revolutionized the field. It is well beyond the scope of the present entry to describe Frege's achievements in this work. But it should be said that one of his most important contributions was to achieve for the first time a clear understanding of the semantic functioning of expressions of generality, such as 'every,' 'some' and so on. This made it possible to understand, again for the first time, how sentences containing multiple expressions of generality, such as 'Every skier loves some mountain,' manage to mean what they do. In a series of papers written in the late 1800s, Frege articulated a novel theory of meaning for languages that was to be very influential. These papers

included "Function and Concept" (1891), "On Concept and Object" (1892) and most famously "On Sense and Reference" (1892).

Frege made a fundamental distinction between expressions that are *unsaturated* or *incomplete* and expressions that are *complete*. The former he called *concept words* (perhaps *concept expressions* would be better) and the latter he called *proper names*. A sentence like:

1. Frege runs.

can be split up into the part that is unsaturated, the concept word 'runs,' and the complete part, the proper name 'Frege.' All expressions, Frege thought, are associated with a *sense* and a *reference*. These both have some claim to be called *the meaning* of the expression in question, and so it is probably best to think of Frege as claiming that there are two components to the meaning of an expression. The referent of an expression can be thought of as the thing in the world the expression stands for. Thus, the referent of the proper name 'Frege' is Frege himself. And the referent of the concept word 'runs' is a *concept*, which Frege took to be a function from an object to a truth value. So the concept 'runs' refers to maps an object *o* to the truth value true iff *o* runs. Otherwise, it maps the object to false. By contrast the sense of an expression Frege thought of as a way or mode in which the referent of the expression is presented. So perhaps Frege can be "presented" as the author of *Begriffsschrift*. Then the sense of the name 'Frege' is the descriptive condition *the author of Begriffsschrift*. It is perhaps more difficult to think of senses of concept words, but it helps to think of them as descriptive conditions that present the concept that is the referent in a certain way.

Now Frege thought that the sense of an expression determines its referent. So the sense of 'Frege' is a mode of presentation of Frege, a descriptive condition that Frege uniquely satisfies in virtue of which he is the referent of 'Frege.' Further, in understanding a linguistic expression, a competent speaker grasps its sense and realizes that it is the sense of the expression.

Of course complex linguistic expressions, such as 1 above, also have senses and references. Frege held that the sense of a complex expression is determined by the senses of its parts and how those parts are combined. (Principles of this general sort are called *principles of compositionality*, and so it could be said that Frege held a principle of compositionality for senses.) Indeed, Frege seems to have held the stronger view that the sense of a complex expression is literally built out of the senses of its parts. In the case of 1, its sense is the result of combining the sense of

'runs' and of 'Frege.' Frege believed that just as the expression 'runs' is unsaturated, so its sense too must be unsaturated or in need of completion. The sense of 'Frege,' by contrast, like the expression itself, is whole and complete (not in need of saturation). The sense of 1 is the result of the whole sense of 'Frege' saturating or completing the incomplete/unsaturated sense of 'runs.' It is the unsaturated sense of 'runs' that holds the sense of 1 together, and this is true generally for Frege. Frege called the sense of a declarative sentence like 1 a *thought*. Thus in "On Concept and Object" (p. 193) Frege writes:

For not all the parts of a thought can be complete; at least one must be unsaturated or predicative; otherwise they would not hold together.

Similarly, Frege held that the reference of a complex expression is determined by the references of its parts and how they are put together (i.e. he held a principle of compositionality for referents). In the case of 1, the referent is determined by taking the object that is the referent of 'Frege' and making it the argument of the function that 'runs' refers to. This function maps objects to the True or the False depending on whether they run or not. Thus, the result of making this object the argument of this function is either the True or the False. And whichever of these is the result of making the object the argument of the function is the referent of 1. So sentences have thoughts as senses and truth values (the True; the False) as referents.

Concerning Frege's account of sentences containing *quantifiers*, expressions of generality such as 'every,' 'some' etc., consider the sentence

2. Every student runs.

The words 'student' and 'runs' are both concept words. Thus they have unsaturated senses and refer to concepts: functions from object to truth values. Now Frege thought that a word like 'every' was doubly unsaturated. To form a whole/complete expression from it, it needs to be supplemented with two concept words ('student' and 'runs' in 2). The sense of 'every' is also doubly unsaturated. Thus the sense of 2 is a thought, a complete sense, that is the result of the senses of 'student' and 'runs' both saturating the doubly unsaturated sense of 'every' (in a certain order). By contrast, the referent of 'every' must be something that takes two concepts (those referred to by 'student' and 'runs' in 2) and yields a referent for the sentence. But as we have seen, a sentence's referent is a truth value. Thus the referent of 'every' must take two concepts and return a truth value. That is, its referent is a function from a pair of concepts to a truth value. In

essence, 'every' refers to a function that maps the concepts A and B (in that order) to the True iff every object that A maps to the true, B maps to the true (i.e. iff every object that falls under A falls under B).

Above it was mentioned that Frege thought that the referent of a complex expression was a function of the referents of its parts and how they are combined (compositionality of reference). Some examples seem to show that this is incorrect. Consider the following:

3. Chris believes that snow is white.

3a. Chris believes that Mt. Whitney is more than 14,000 feet high.

These sentences may well have different referents, that is, truth values. But the embedded sentences ('snow is white'; 'Mt. Whitney is more than 14,000 feet high') have the same referents (the True) and the other parts of the sentences have the same referents as well. But then it would seem that compositionality of reference would require that 3 and 3a have the same reference/truth value. Frege famously gets out of this apparent problem by claiming that 'believes' has the effect of shifting the referents of expressions embedded with respect to it. In 3 and 3a, the shifted referents of the embedded sentences are their usual senses. So in these environments, the sentences have different referents because they express different thoughts outside of contexts involving 'believes' and related devices.

Frege's doctrine of sense and reference constitutes a semantical theory of languages, because it claims that the meanings of linguistic expressions have these two components, and it gives an account of what the senses and referents of different kinds of linguistic expressions are.

Shortly after Frege had worked out his semantical theory of sense and reference, the English philosopher and mathematician Bertrand Russell was working out a theory of the meanings, or information contents of sentences. While Frege had held that the thought expressed by a sentence, which captures the information the sentence encodes, consisted of senses, Russell (1903) held that the information encoded by sentences were *propositions*, where the constituents of propositions, far from being Fregean senses, where roughly (and for the most part) the things the propositions is about. Thus, whereas Frege held that 1 expressed a thought containing a mode of presentation of Frege and a mode of presentation of the concept of running, Russell held that the proposition expressed by 1 contained Frege himself and the concept of running (though Russell thought of concepts differently from the way Frege did). This contrast has more

than historical significance, because current semanticists are classified as Fregean or Russellian depending on whether they hold that the information contents of sentences contain the things those information contents are about (objects, properties and relations—Russellian) or modes of presentation of the things those information contents are about (Fregean).

In the early part of the twentieth century, the philosophical movement known as *Logical Positivism* achieved dominance, especially among logically minded philosophers who might have been interested in semantics. The Positivists thought that much of traditional philosophy was literally nonsense. They applied the (pejorative) term “*metaphysics*” to what they viewed as such philosophical nonsense. The Positivists, and especially Rudolf Carnap, developed accounts of meaning according to which much of what had been written by philosophers was literally meaningless. The earliest and crudest Positivist account of meaning was formulated by Carnap (1932). On this view, the meaning of a word was given by first specifying the simplest sentence in which it could occur (its *elementary sentence*). Next, it must be stated how the word’s elementary sentence could be verified. Any word not satisfying these two conditions was therefore meaningless. Carnap held that many words used in traditional philosophy failed to meet these conditions and so were meaningless.

Carnap called philosophical statements (sentences) that on analysis fail to be meaningful *pseudo-statements*. Some philosophical statements are pseudo-statements, according to Carnap, because they contain meaningless terms as just described. But Carnap thought that there is another class of philosophical pseudo-statements. These are statements that are literally not well formed (Carnap gives Heidegger’s “We know the nothing.” as an example).

The downfall of the Positivist’s theory of meaning was that it appeared to rule out certain scientifically important statements as meaningless. This was unacceptable to the Positivists themselves, who were self-consciously very scientifically minded. Carnap heroically altered and refined the Positivists account of meaningfulness, but difficulties remained. Hempel (1950) is a good source for these developments.

At about the same time Carnap was formulating the Positivists’ account of meaning, the Polish logician Alfred Tarski was involved in investigations that would change forever both logic and semantics. It had long been thought that meaning and truth were somehow intimately connected. Indeed, some remarks of Wittgenstein’s in his *Tractatus Logico-Philosophicus* (“4.024. To

understand a proposition means to know what is the case, if it is true.”) had led many to believe that the meaning of a sentence was given by the conditions under which it would be true and false. However, the Positivists had been wary of the notion of truth. It seemed to them a dangerously metaphysical notion, (which is why they “replaced” talk of *truth* with talk of *being verified*).

Against this background, Tarski showed that truth (‘true sentence’) could be rigorously defined for a variety of formal languages (languages, growing out of Frege’s work in logic, explicitly formulated for the purpose of pursuing research in logic or to be used to precisely express mathematical or scientific theories). Though earlier papers in Polish and German contained the essential ideas, it was Tarski (1935) that alerted the philosophical world to Tarski’s important new results.

Tarski himself despaired of giving a definition of *true sentence of English* (or any other naturally occurring language). He thought that the fact that such languages contain the means for talking about expressions of that very language and their semantic features (so English contains expressions like ‘true sentence,’ ‘denotes,’ ‘names,’ etc.) meant that paradoxes, such as the paradox of the liar, are formulable in such languages. In turn, Tarski thought that this meant that such languages were logically inconsistent and hence that there could be no correct definition of ‘true sentence’ for such languages.

Nonetheless, Tarski’s work made the notion of truth once again philosophically and scientifically respectable. And it introduced the idea that an important element, perhaps the sole element, in providing a semantics for a language was to provide a rigorous assignment to sentences of the language the conditions under which they are true. (Tarski’s 1935 paper for the most part gave definitions of *true sentence* for languages with fixed interpretations. The now more familiar notion of *true sentence with respect to a model* was introduced later. See Hodges [2001] for details.)

Carnap’s *Meaning and Necessity* (1947) is arguably the first work that contemporary semanticists would recognize as a work in what is now considered to be semantics. Following Tarski, Carnap distinguishes the languages under study and for which he gives a semantics, *object languages*, from the languages in which the semantics for the object languages are stated, *metalanguages*. The object languages Carnap primarily considers are a standard first order language ( $S_1$ ), the result from adding ‘N’ (“a sign for logical necessity”) to that language ( $S_2$ ), and ordinary English. Carnap does not give detailed descriptions of any of these languages, noting that the book

“... is intended not so much to carry out exact analyses of exactly constructed systems as to state informally some considerations aimed at the discovery of concepts and methods suitable for semantical analysis” (p. 8).

The heart of Carnap’s semantics for these languages is given by *rules of designation* for predicates and individual constants, *rules of truth* for sentences and *rules of ranges* for sentences. The rules of designation state the meanings of the predicates and individual constants using English as the metalanguage. So we have (p. 4):

‘s’ is a symbolic translation of ‘Walter Scott’  
 ‘Bx’—‘x is a biped’

The rules of truth simply provide a Tarski style definition of truth for sentences of the language, (the definition assumes fixed meanings given by the rules of designation for predicates and individual constants). In order to specify the rules of range, Carnap introduces the notion of a *state-description*. For a language, say  $S_1$ , a state description in  $S_1$  is a set that contains for every atomic sentence of  $S_1$ , either it or its negation, but not both; and it contains no other sentences. Carnap comments (p. 9):

... it [a state-description in  $S_1$ ] obviously gives a complete description of a possible state of the universe of individuals with respect to all properties and relation S expressed by predicates of the system. Thus the state-descriptions represent Leibniz’ possible worlds or Wittgenstein’s possible states of affairs.

Next Carnap gives a recursive characterization of a *sentence holding in a state-description*. An atomic sentence holds in a state-description iff it is a member of it. A disjunction holds in it iff one of its disjuncts holds in it, etc. The characterization of *holding in a state description* is designed to formally capture the intuitive idea of the sentence being true if the possible world represented by the state-description obtained (i.e. if all the sentences belonging to the state-description were true). Given a sentence S, Carnap calls the class of state-descriptions in which S holds its *range*. Thus the clauses in the characterization of *holding in a state-description* Carnap calls *rules of ranges*. Regarding these rules of ranges, Carnap writes (p. 9–10):

By determining the ranges, they give, together with the rules of designation for the predicates and the individual constants ..., an *interpretation* for all sentences of  $S_1$ , since to know the meaning of a sentence is to know in which of the possible cases it would be true and in which not, as Wittgenstein has pointed out.

Thus, Carnap regards the rules of ranges together with the rules of designation as giving the meaning of the sentences of  $S_1$  (the connection with truth and the rules of truth is that there is one state-description that describes the actual world, and a sentence is true iff it holds in that state-description).

Using these resources, Carnap defines his well known *L concepts*. We here concentrate on *L-truth* and *L-equivalence*. Before getting to that, we must say something about Carnap’s notion of *explication*. Carnap believed that one of the main tasks for philosophers was to take a “vague or not quite exact” concept, and replace it by a more exact concept that one had clearly characterized. This new concept, called by Carnap the *explicatum* of the old concept, was intended to be used to do the work the old concept was used to do. Carnap thought that the notion of *L-truth* was the explicatum of the vague notions of “logical or necessary or analytic truth” (p. 10).

A sentence is L-true in a semantical system (e.g.  $S_1$ ) iff it holds in every state description in that system. Carnap regarded this as a precise characterization of Leibniz’s idea that necessary or analytic or logical truths hold in all possible worlds. Next, Carnap defines the notion of *L-equivalence* for sentences, predicates and individual constants. Effectively, two names, predicates or sentences are L-equivalent (in a semantical system—e.g.  $S_1$ ) iff they have the same extension at every state-description in that system, (so L-equivalent names must name the same individual at every state description, L-equivalent predicates must be true of the same individuals at every state description, etc.).

The importance of Carnap’s notion of *L-equivalence* is that he uses it to sketch a semantics for belief ascriptions. In order to do this, Carnap extends his notion of *L-equivalence* in several ways. First, he extends it so that expressions of different “semantical systems” (roughly, formal languages) may be L-equivalent (in effect, expressions e of system 1 and e’ of system 2 are L-equivalent just in case the semantical rules of the two systems together suffice to show that the expressions have the same extension, p. 57). Second, he extends the notion of *L-equivalence* to apply to sentential connectives, variables (they are L-equivalent iff they have the same range of values) and to quantifiers (they are L-equivalent iff they are quantifiers of the same sort [universal, existential] whose variables are L-equivalent, p. 58). Third, he defines what it is for two expressions of the same or different semantical systems (again, roughly formal languages) to be *intensionally isomorphic*. Roughly, expressions are intensionally isomorphic just in case they are built up in the same

way out L-equivalent parts. With these tools in hand, Carnap writes (p. 61–62):

It seems that the sentence ‘John believes that D’ in S [a fragment of English—see p. 53] can be interpreted by the following semantical sentence:

15-1. ‘There is a sentence  $\mathfrak{S}_1$  in the semantical system S’ such that (a)  $\mathfrak{S}_1$  is intensionally isomorphic to ‘D’ and (b) John is disposed to an affirmative response to  $\mathfrak{S}_1$

Though Carnap’s semantics for belief ascriptions was criticized by Alonzo Church (1950), many philosophers were influenced by Carnap’s idea that the objects of belief are structured entities built up in the same way out of entities with the same intensions. See, for example, Lewis (1970).

The final important feature of *Meaning and Necessity* was its semantic treatment of modality. Carnap begins his discussion of modality by mentioning the work of C. I. Lewis (presumably he had in mind especially Lewis and Langford [1932]) in constructing various systems of modal logic. As mentioned above, Carnap considered as an object of semantical investigation a language that was the first order predicate logic ( $S_1$ ) supplemented with the sign ‘N’ “for logical necessity.” He called the resulting language  $S_2$ . Syntactically, prefixing ‘N’ to a matrix (either a sentence or a formula with free variables) results in a matrix. A detailed discussion of Carnap’s semantics for this modal language would go beyond the scope of the present entry. However, a couple points are worth making. First, if we just consider the case in which ‘N’ fronts a *sentence* (formula with no free variables)  $\phi$ , to get the rules of range for  $S_2$  we would simply add to the rules of range of  $S_1$  the following:

N( $\phi$ ) holds in every state-description if  $\phi$  holds in every state description; otherwise N( $\phi$ ) holds in no state-description.

This is a consequence of Carnap’s idea that ‘N’ is the sign for logical necessity, and the notion of L-truth is the explicatum of the vague notion of logical necessity. Thus a sentence fronted by ‘N’ should hold at a state description iff the sentence it embeds holds at every state-description. But then if the sentence fronted by ‘N’ holds at a state-description, it holds at every state-description. Thus, the above.

But of course since ‘N’ could front a matrix with free variables, one could then attach a quantifier to the result. Letting ‘.u.’ be a matrix containing the variable ‘u’ free, we get things like

(u)N(.u..)

That is, we get quantifying into the sign ‘N’ for logical necessity. However, Carnap’s treatment here results in the above being equivalent to (indeed, L-equivalent to)

N(u)(.u..).

The important point, however, is that Carnap had sketched a semantics for quantified modal logic.

Though virtually all of the crucial analyses and explications in *Meaning and Necessity* were eventually significantly modified or rejected (the explication of “logical necessity” by the notion of L-truth, understood in terms of holding at all state-descriptions; the treatment of ‘N’, the sign of “logical necessity”; and the semantics for belief ascriptions), the work was nonetheless very important in the development of semantics. It provided a glimpse of how to use techniques from logic to systematically assign semantic values to sentences of languages, and began the project of providing a rigorous semantics for recalcitrant constructions like sentences containing modal elements and verbs of propositional attitude.

In the 1950s and early 1960s Carnap’s ideas on the semantic treatment of modal logic were refined and improved upon. The result was the now familiar “Kripke style” semantics for modal logic. Kripke’s formulations will be discussed here, but it is important to understand that similar ideas were in the air (see Hintikka [1961], Kanger [1957], and Montague [1960a]). Though these works were in the first instance works in logic, as we will see, they had a profound effect on people who were beginning to think about formal semantics for natural languages.

We will concern ourselves with the specific formulations in Kripke (1963). What follows will be of necessity slightly technical. The reader who is not interested in such things can skip to the end of the technical discussion for informal remarks. Assume that we have a standard first order logic with sentential connectives  $\sim, \&$  and  $\square$  (the first and third one-place, the second two-place), individual variables (with or without subscripts)  $x, y, z, \dots$ ; n-place predicates  $P^n, Q^n, \dots$  (0 place predicate letters are *propositional variables*), and universal quantifier (for any variable  $x_i$ , ( $x_i$ )). A *model structure* is a triple  $\langle G, K, R \rangle$ , where K is a set,  $G \in K$  and R is a *reflexive relation* on K (i.e. for all  $H \in K$ ,  $H R H$ ). Intuitively, G is the “actual world” and the members of K are all the possible worlds. R is a relation between worlds and is usually now called the *accessibility relation*. Intuitively, if  $HR H'$  ( $H'$  is accessible from H), then what is *true* in  $H'$  is *possible* in H.

Again intuitively, the worlds accessible from a given world are those that are possible relative to it.

Putting conditions on  $R$  gives one model structures appropriate to different modal logics. If  $R$  is merely reflexive, as required above, we get an  $M$  model structure. If  $R$  is reflexive and *transitive* (i.e. for any  $H, H', H'' \in K$ , if  $H R H'$  and  $H' R H''$ , then  $H R H''$ ), we get an  $S4$  model structure. Finally, if  $R$  is reflexive, transitive and *symmetric* (i.e. for any  $H, H' \in K$ , if  $H R H'$ , then  $H' R H$ ), we get an  $S5$  model structure. (It should be recalled that for Carnap, state-descriptions, which represented possible worlds, were each accessible for every other—in effect because there was no accessibility relation between state-descriptions; thus translated into the present framework Carnap’s “models” would be  $S5$  models. Also, in Kripke’s semantics, possible worlds (members of  $K$ ) are primitive; in Carnap’s, of course, they are explicated as state descriptions.) A *quantificational model structure* is a model structure  $\langle G, K, R \rangle$  together with a function  $\psi$  that assigns to every  $H$  in  $K$  a set of individuals: the *domain* of  $H$ . Intuitively this is the set of individuals existing in the possible world  $H$ . Of course, this allows different worlds (members of  $K$ ) to have different domains of individuals. This formally captures the intuitive idea that some individuals that exist might not have, and that there might have been individuals that there aren’t.

Given a quantificational model structure, consider the set  $U$  which is the union of  $\psi(H)$  for all  $H$  in  $K$ . Intuitively, this is the set of all possible individuals (i.e. the set  $U$  of individuals such that any individual in the domain of any world is in  $U$ ). Then  $U^n$  is the set of all  $n$ -tuples whose elements are in  $U$ . A *quantificational model* on a quantificational model structure  $\langle G, K, R \rangle$  is a function  $\varphi$  that maps a zero-place predicate and a member of  $K$  to  $T$  or  $F$ ; and for  $n > 0$ , an  $n$ -place predicate and a member of  $K$  to a subset of  $U^n$ . We extend  $\varphi$  by induction to assign truth values to all formula/world pairs *relative to a function assigning members of  $U$  to variables*:

1. *Propositional Variable*: Let  $f$  be a function assigning elements of  $U$  to all individual variables. Let  $P$  be a propositional variable. Then for any  $H$  in  $K$ ,  $\varphi(P, H) = T$  relative to  $f$  iff  $\varphi(P, H) = T$ ; otherwise  $\varphi(P, H) = F$  relative to  $f$ .

2. *Atomic*: Let  $f$  be as in 1. For any  $H$  in  $K$ ,  $\varphi(P^n x_1, \dots, x_n, H) = T$  relative to  $f$  iff  $\langle f(x_1), \dots, f(x_n) \rangle \in \varphi(P^n, H)$ ; otherwise  $\varphi(P^n x_1, \dots, x_n, H) = F$  relative to  $f$ .

(Note that 2 allows that an atomic formula can have a truth value at a world relative to an assignment to its variables, where some or all of its variables get assigned things

not in the domain of the world, since  $f$  assigns elements of  $U$  to free variables; and  $\varphi$  assigns subsets of  $U^n$  to  $P^n$ !)

3. *Truth functional connectives*: Let  $f$  be as in 1. Let  $A$  and  $B$  be formulae. For any  $H$  in  $K$ ,  $\varphi(A \& B, H) = T$  relative to  $f$  iff  $\varphi(A, H) = T$  relative to  $f$  and  $\varphi(B, H) = T$  relative to  $f$ ; otherwise  $\varphi(A \& B, H) = F$  relative to  $f$ . (Similarly for  $\sim$ )

4. *Modal operator*: Let  $f$  be as in 1. Let  $A$  be a formula.  $\varphi(\Box A, H) = T$  relative to  $f$  iff  $\varphi(A, H') = T$  relative to  $f$  for all  $H' \in K$  such that  $H R H'$ ; otherwise  $\varphi(\Box A, H) = F$  relative to  $f$ .

(Note that according to 4, whether a formula  $\Box A$  is true at a world (relative to  $f$ ) depends only on whether  $A$  is true at all worlds *accessible* from the original world.)

5. *Quantifier*: Let  $f$  be as in 1. Let  $A(x, y_1, \dots, y_n)$  be a formula containing only the free variables  $x, y_1, \dots, y_n$ . For any  $H$  in  $K$ , and any function  $g$  (assigning elements of  $U$  to free variables), suppose  $\varphi(A(x, y_1, \dots, y_n), H)$  relative to  $g$  is defined. Then  $\varphi((x) A(x, y_1, \dots, y_n), H) = T$  relative to  $f$  iff for every  $f'$  such that  $f'(x) \in \psi(H)$  and  $f'$  differs from  $f$  at most in that  $f'(x)$  is not  $f(x)$ ,  $\varphi(A(x, y_1, \dots, y_n), H) = T$  relative to  $f'$ ; otherwise,  $\varphi((x) A(x, y_1, \dots, y_n), H) = F$  relative to  $f$ .

(As Kripke notes, that in 5 we consider only functions  $f'$  such that  $f'(x) \in \psi(H)$  means that quantifiers range over only the objects that exist at the world where the quantified sentence is being evaluated.)

Now having gone through Kripke’s semantics for quantified modal logic in some detail, let us step back and ask why it was important in terms of thinking of the semantics of natural language. People like Richard Montague, who we will discuss below, were clearly influenced in their thinking about the semantics of natural language by Kripke’s semantics for modal logic, (recall too that Montague [1960a] itself contained ideas related to Kripke’s). Since at least Carnap’s *Meaning and Necessity* (and perhaps before), philosophers had thought of sentences as semantically associated with *propositions* and of  $n$ -place predicates as semantically associated with  $n$ -place *relations* (properties being one-place relations). Further, they had thought of these propositions and relations as determining truth values and extensions for the sentences and predicates expressing them relative to a “possible world” (which, of course, Carnap represented by a state description).

Now in Montague (1960b), it is suggested that an  $n$ -place relation just is a function from possible worlds to a set of  $n$ -tuples (intuitively, the set of  $n$ -tuples whose ele-



ments stand in the relation in question from the standpoint of the world in question); and that a proposition just is a function from possible worlds to truth values. Generalizing these ideas leads straightforwardly to possible worlds semantics for natural languages discussed below. Further, Montague claims this way of understanding relations and propositions (which Montague calls *predicates*, one-place predicates, then, are properties; and zero-place predicates are propositions) is to be found for the first time in Kripke (1963). This, in turn, means that at least Montague saw the seeds of possible worlds semantics for natural languages in Kripke (1963).

This initially seems at least a little bit strange, since nowhere in Kripke (1963) does one find the identification of propositions with functions from possible worlds to truth values or relations with functions from possible worlds to sets of  $n$ -tuples. However, it is easy to see why a logician like Montague would see those ideas in Kripke (1963). Consider again a model on a quantificational model structure, forgetting for the moment about functions  $f$  that are assignments to free variables and that the domains of members of  $K$  can vary, (essentially, this means we are considering a model on a *propositional* model structure). A *model*  $\varphi$  on a  $(M/S_4/S_5)$  model structure  $\langle G, K, R \rangle$  assigns to a propositional variable (a zero-place predicate—an atomic formula without any variables) and a member of  $K$  either T or F. Now consider a particular propositional variable  $P$ . Consider the function  $f_p$  defined as follows:

For any  $H$  in  $K$ ,  $f_p(H) = T$  iff  $\varphi(P, H) = T$ ; otherwise  
 $f_p(H) = F$

$f_p$  is a function from worlds to truth values and so can be thought of *a la* Montague as the proposition expressed by  $P$  (in the model  $\varphi$  on the model structure  $\langle G, K, R \rangle$ )! That is, propositions, understood as functions from worlds to truth values, are trivially definable using Kripke's models. Similar remarks apply to  $n$ -place relations, understood as functions from possible worlds to sets of  $n$ -tuples of individuals. It seems likely that this is why a logician like Montague would take Kripke to have introduced them. Montague, after making the attribution to Kripke, does add (p.154): "... Kripke employs, however, a different terminology and has in mind somewhat different objectives."

These functions from worlds to truth values or sets of  $n$ -tuples are now generally called *intensions*. Their values at a world (truth values; sets of  $n$ -tuples) are generally called *extensions* (at worlds). The idea that the primary job of semantics is to assign to expressions of natural lan-

guages intensions and extensions of the appropriate sort very much took hold in the wake of work by Kripke and others in the semantics of modal logic.

With the resources Kripke and others had made available in hand, researchers thinking about the semantics of natural languages eagerly made use of them. Thus the late 1960s and early 1970s saw dizzying progress in natural language semantics as the techniques for modal logic were applied. Two works from that era that particularly capture the spirit of the times are Lewis (1970) and Montague (1973). The latter will be discussed here, since it is probably the most sophisticated and influential of the works of that period. The particular semantic phenomena Montague was concerned to understand were the workings of verbs of propositional attitudes like 'believes,' the workings of intensional verbs like 'worships' and related phenomena (see p. 248 where Montague lists some of his concerns).

We saw above that both Frege and Carnap were also concerned with understanding the semantics of verbs like 'believes.' We are now in a position to say more about why such expressions attract the attention of semanticists. Consider the expression 'It is not the case' in sentences like

4. It is not the case that snow is white.

4a. It is not the case that Mt. Whitney is more than 14,000 feet high.

Whether a sentence fronted by 'It is not the case' is true or false depends only on the extension/truth value of the embedded sentence. Since both the embedded sentences are true, 4 and 4a are both false. Let's put this by saying that 'It is not the case that' creates *extensional contexts*. As we saw above, 'believes' doesn't create extensional contexts. 3 and 3a can differ in truth value even though the embedded sentences are both true. Let's say that 'believes' creates *nonextensional contexts*. The same is true of 'Necessarily.' The following differ in truth value even though the embedded sentences have the same extensions/truth values:

5. Necessarily, everything is identical to itself.

5a. Necessarily, Aristotle is a philosopher.

Finally, intensional verbs like 'worship' exhibit similar behavior and we could extend our characterization of creating nonextensional contexts so as to include such verbs. For even though 'Samuel Clemens' and 'Mark Twain' have the same extension (a certain individual), the following two sentences apparently may differ in extension/truth value:

6. Lori worships Samuel Clemens.

6a. Lori worships Mark Twain.

Now semanticists have been puzzled as to how to think of the semantics of expressions that create nonextensional contexts. But the work of Carnap and Kripke suggested the way to understand ‘Necessarily.’ In particular,

Necessarily S is true at a world  $w$  just in case the intension of S maps every world (accessible from  $w$ ) to true.

In other words, whereas ‘It is not the case’ looks at the extension of the sentence it embeds to determine whether the entire sentence containing it is true, ‘Necessarily’ looks at the intension of the sentence it embeds to determine whether the entire sentence containing it is true. And given Kripke’s semantics, intensions were well defined, respectable entities: functions from worlds to extensions. This made it appear to many that a semantics that assigned intensions to expressions could treat all expressions creating nonextensional contexts. Certainly, Montague had a version of this view.

As indicated above, Montague (1973) wanted to provide semantic treatments of verbs of propositional attitude such as ‘believes,’ intensional verbs such as ‘worships,’ and other phenomena. We will concentrate on these phenomena as well as Montague’s treatment of quantification. Montague (1973) provides a syntax for a fragment of English. The fragment includes common nouns (‘woman’; ‘unicorn’), intransitive verbs (including ‘run’ and ‘rise’), transitive verbs, (including both intensional transitives and “normal” transitive verbs like ‘love’), ordinary names and pronouns, adverbs (including ‘rapidly’ and ‘allegedly’), prepositions, verbs of propositional attitude and modal sentence adverbs (“adsentences”—‘necessarily’). The fragment allows the formation of relative clauses (though they employ the somewhat stilted ‘such that,’ so that we get things like ‘man such that he loves Mary’) and so complex noun phrases, as well as prepositional phrases and quantifier phrases (‘Every woman such that she loves John’). Thus, Montague’s syntactic fragment includes sentences like:

7. Every man loves a woman such that she loves him.

8. John seeks a unicorn.

9. John talks about a unicorn.

10. Mary believes that John finds a unicorn.

11. Mary believes that John finds a unicorn and he eats it.

It should be noted that many sentences of Montague’s fragment had non-trivially different syntactic analyses: that is, distinct syntactic analyses that are interpreted differently semantically. So, for example, 8 above has an analysis on which ‘a unicorn’ is the constituent last added to the sentence and an analysis on which ‘John’ is the last constituent added. The latter has an interpretation on which it may be true even if there are no unicorns and so John is seeking no particular one. The former requires John to be seeking a particular unicorn. Thus, it is really syntactic analyses of sentences, and not the sentences themselves, that get semantic interpretations.

The next aspect of Montague’s semantic treatment of his fragment of English is his intensional logic. Montague’s intensional logic is typed. In particular,  $e$  and  $t$  are the basic types; and whenever  $a$  and  $b$  are types,  $\langle a, b \rangle$  is a type. Finally, for any type  $a$ ,  $\langle s, a \rangle$  is a type. For each type, there will be both constants and variables of that type (and hence quantifiers of that type). The key to understanding the syntactic interactions of the expressions of various types is to know that if  $\alpha$  is of type  $\langle a, b \rangle$  and  $\beta$  is of type  $a$ , then  $\alpha(\beta)$  is of type  $b$ . Interpretations assign expressions of the logic various denotations (relative to an assignment of values to variables). Expressions of type  $e$  get assigned individuals (possible individuals); expressions of type  $t$  get assigned truth values. Expressions of type  $\langle a, b \rangle$  get assigned as denotations functions from denotations of type  $a$  to denotations of type  $b$ . Finally, expressions of type  $\langle s, a \rangle$  get assigned functions from a world/time pair to a denotation of type  $a$  (“an intension of a type  $a$  expression”). To take some examples, expressions of type  $\langle e, t \rangle$  get assigned functions from individuals to truth values (the denotations can alternatively be thought of as sets of individuals: those that get assigned to true). Expressions of type  $\langle s, e \rangle$  are assigned functions from world/time pairs to individuals. Such functions Montague called *individual concepts*. Expressions of type  $\langle \langle s, e \rangle, t \rangle$  are assigned functions from individual concepts to truth values (alternatively, sets of individual concepts). Expressions of type  $\langle s, t \rangle$  are assigned functions from world/time pairs to truth values. As indicated above, Montague thought of these as propositions.

The way Montague provided a semantic interpretation of his syntactic fragment of English was to provide an algorithm for translating English sentences (really, syntactic analyses of English sentences) into his intensional logic. Then the interpretation of the English sentences was given by the interpretation of its translation in intensional logic. Recall again that sentences like 8 above can be true even if there are no unicorns. Thus, a verb like

‘seeks’ could not have as its denotation (really, its translation into intensional logic could not have as its denotation) a relation between individuals (or a function from individuals to a function from individuals to truth values).

In order to get the proper results, Montague decided to assign to common nouns and intransitive verbs as their denotations sets of individual concepts rather than sets of individuals. Verbs like ‘believes’ have as their denotations functions from propositions to sets of individual concepts. Since individual concepts essentially function as individuals in Montague’s semantics (recall that common nouns like ‘man’ have as denotations sets of individual concepts), this treatment essentially amounts to holding that verbs of propositional attitude denote relations between individuals and propositions. Quantifiers such as ‘Every man’ denote sets of properties of individual concepts (functions from world/time pairs to sets of individual concepts). Roughly, ‘Every man walks’ is true at a world and time  $\langle w, t \rangle$  just in case the property of individual concepts that determines the correct set of individual concepts denoted by ‘walks’ at every world and time is in the set of properties of individual concepts denoted by ‘Every man’ at  $\langle w, t \rangle$ . ‘Necessarily’ denotes at a world/time  $\langle w, t \rangle$  a set of propositions: those that are necessary at  $\langle w, t \rangle$ .

Finally, a transitive verb denotes a function from properties of properties of individual concepts (denotations of expressions of type  $\langle s, \langle \langle s, \langle \langle s, e \rangle, t \rangle \rangle, t \rangle \rangle$ )—functions from world/time pairs to sets of properties of individual concepts) to sets of individual concepts. Again, recalling that individual concepts essentially stand in for individuals in Montague’s framework, this means that transitive verbs in effect denote relations between individuals and properties of properties of individuals. Note that this means that for  $\delta$  to be true at a world/time pair  $\langle w, t \rangle$  is for John to stand in a relation to the property of being a property possessed by a unicorn. This can be the case even if there are no unicorns.

Montague chose to treat all expressions of a given syntactic category the same way semantically. This means that transitive verbs like ‘loves’ get the odd denotation required by ‘seeks’ to get  $\delta$  right. But don’t we want ‘John loves Mary’ to be true at world/time pair iff the individual John stands in a relation to the individual Mary? Surely this shouldn’t require instead that John stands in a relation to the property of being a property possessed by Mary. Where’s the love (between individuals)? Montague essentially requires interpretations to make true meaning postulates for “ordinary” verbs like ‘loves,’ and these end

up insuring that ‘John loves Mary’ is true at  $\langle w, t \rangle$  iff John and Mary themselves are properly related.

Montague’s semantic account here was very influential. He showed that the resources Kripke and others developed for the semantics of modal logic could be rigorously applied to natural languages, and arguably treat such recalcitrant expressions as ‘believes,’ ‘necessarily,’ and ‘seeks.’ Montague’s basic approach was picked up by many philosophers and linguists and much work in semantics through the 1980s and beyond was conducted in this framework. Indeed, much work is still done in this and closely related frameworks.

At about the same time Montague was doing his pioneering work on formal semantics for natural languages, Donald Davidson was developing a very different approach to semantics. Davidson (1967) begins with the idea that a theory of meaning for a natural language must specify how the meaning of a sentence is determined by the meanings of the words in it, and presumably how they are combined (in other writings, Davidson puts the point by saying that the meaning of sentence must be a function of a finite number of features of the sentence—presumably, one is its syntax). Davidson thought that only a theory of this sort could provide an explanation of the fact that on the basis of mastering a finite vocabulary and a finite number of syntactic rules, we are able to understand a potentially infinite number of sentences. More specifically, Davidson thought a theory of meaning should comprise an axiomatized theory, with a finite number of axioms, that entails as theorems (an infinite number of) statements specifying the meaning of each sentence of the language. Davidson thought that grasping such a theory would allow one to understand all the sentences of the language. Further, as suggested above, such a theory would explain how creatures like us are capable of understanding an infinite number of sentences. It would only require us to grasp the axioms of the theory of meaning, which are finite in number.

It might be thought that the theorems of a theory of meaning of the sort discussed would be all true sentences of the form ‘*s* means that *p*,’ where ‘*s*’ is replaced by a structural description of a sentence of the language and ‘*m*’ is replaced by a term referring to a meaning. Further, it might be thought that a theory would have such theorems in part by assigning meanings to the basic expressions of the language (such assignments being made by axioms). However, Davidson thinks that we have not a clue as to how to construct such a theory, mainly because we have no idea how the alleged meanings of simpler expressions combine to yield the meanings of complex

expression of which they are parts. Thus, Davidson concludes, postulating meanings of expressions gets us nowhere in actually giving a theory of meaning for a language.

Davidson's counterproposal as to what a theory of meaning should be like is radical. A theory of meaning must be a finite number of axioms that entail for every sentence of the language a true sentence of the form 's is true iff p', where 's' is replaced by some sort of description of a sentence whose theory of meaning we are giving, and 'p' is replaced by some sentence. Henceforth, we will call such sentences *T-sentences*. Recalling our discussion of Tarski, the language we are giving a theory of meaning for is the object language and the theory of meaning is given in the metalanguage. Thus, the formulation just given requires the metalanguage to have some sort of (presumably standardized) description of each sentence of the object language (to replace 's'); if we imagine 'p' to be replaced by the very sentence that what replaces 's' describes (as Davidson sometimes supposes) the metalanguage must also contain the sentences of the object language. In short, Davidson held that to give a theory of meaning for a language is to give a Tarski-style truth definition for it.

Tarski thought that a condition of adequacy for a theory of truth for a (in his case, formal) language L was that the theory has as consequences all sentences of the form 's is true (in L) iff p', where 's' is replaced by a structural description of a sentence of the object language and 'p' is replaced by a translation of it. Here Tarski clearly seemed to think that for one sentence to translate another is for them to share a meaning. However in characterizing what is to replace 'p' in his T-sentences, Davidson cannot require 'p' to be replaced by a translation of the sentence the thing replacing 's' describes, assuming anyway that for one sentence to be a translation of another is for them to share the same meaning. For Davidson eschews meanings. After all, a theory of truth was supposed to *be* a theory of meaning; it would hardly do, then, to appeal to meanings in constructing one's theory of truth. Thus Davidson famously merely requires the T-sentences to be true. But this requirement is very weak, for 'iff' is truth functional in Davidson's T-sentences, and so the sentences require for their truth only that the two sides share a truth value. But then there is nothing in principle yet to prevent having a theory of truth for English that yields not:

12. 'Snow is white' is true (in English) iff snow is white.

but instead

13. 'Snow is white' is true (in English) iff grass is green.

After all, 13 is true! Davidson was aware of this consequence of his view, and explicitly discussed it. He claimed that by itself, the fact that a theory of truth yields 13 as a theorem instead of 12 doesn't cut against it. However, the theory has to get all the other T-sentences coming out true, and Davidson thought it was unlikely that it could do that and yield 13 as a theorem.

Of course, the picture sketched so far needs to be complicated to account for contextually sensitive expressions. It won't do to have as theorems of one's truth theory things such as:

14. 'I am hungry' is true (in English) if I am hungry.

Davidson himself thought that the way to deal with this was to relativize truth to e.g. a speaker and a time (to handle tense). Others have suggested that a theory of truth for a language containing such contextually sensitive words must define truth for utterances of sentences. For example, see Weinstein (1974).

Further complications are required as well. Natural language contains devices not contained in the relatively austere formal languages for which Tarski showed how to define truth. Natural languages contain verbs of propositional attitude ('believes'), non-indicative sentences and other features. Davidson attempted to provide accounts of many such devices in other papers. Davidson (1968) for example takes up verbs of propositional attitude.

One sometimes hears model theoretic approaches to semantics contrasted with those that offer an absolute truth theory. The contrast is illustrated by comparing Montague and Davidson, since each is perhaps the paradigmatic case of one of these approaches. As we saw, Montague gives a semantics for English sentences by associating them with formulae of intensional logic. He then gives a semantics for the formulae of intensional logic. Now the latter includes a definition of truth relative to an interpretation (and other parameters as well). As discussed, expressions of Montague's intensional logic only have denotations (and intensions) relative to interpretations, which are also sometimes called *models*. Roughly, then, a model theoretic semantics is one that defines truth *relative to models or interpretations*. By contrast, as we have seen, Davidson wants a theory of truth simpliciter (actually, truth for L, but truth isn't relativized to models). Thus, Davidson's approach is sometimes called an absolute truth theory approach. I believe it is

fair to say that most semanticists today use a model theoretic approach.

The 1960s and 1970s saw an explosion in the sort of model theoretic semantics pioneered by Montague, Lewis and others. Some of the important developments had to do with evolving notions of an *index of evaluation*. As we saw above, in Montague's intensional logic, expressions are assigned extensions/denotations at world/time pairs (under an interpretation relative to an assignment of values to variables—this will be suppressed in the present discussion for ease of exposition). In particular, formulae are assigned truth values at a pair of a world and time.

Since expressions of Montague's English fragment receive semantic interpretations by being given the interpretation assigned to the expressions of intensional logic they are translated into, exactly similar remarks apply to English expressions and sentences. We shall call these elements at which expressions are assigned extensions (in this case, world/time pairs) *indices*. (Terminology here varies: Montague called these things *points of reference*; Lewis [1970] called them *indices*, which is probably the most common term for them.) It should be obvious why sentences are assigned truth value at worlds. The reason Montague included times in his indices was that his intensional logic included tense operators in order that he could capture the rudimentary behavior of tense in English. Semantically, such operators work by shifting the time element of the index. Thus, where P is a past tense operator,  $\phi$  a formula, w a world and t a time,  $P\phi$  is true at  $\langle w, t \rangle$  iff  $\phi$  is true at  $\langle w, t' \rangle$  for some  $t'$  prior to t. Similarly, modal operators shift the world element of the index: Necessarily  $\phi$  is true at  $\langle w, t \rangle$  iff  $\phi$  is true at  $\langle w', t \rangle$  for all  $w'$ .

So the truth values of formulae of Montague's intensional logic, and so of the English sentences they translate, depend on (or vary with) both a world and a time. Of course, it was noticed that the truth values of some English sentences vary with other features as well, such as who is speaking (if the sentence contains 'I'); who is being addressed (if the sentence contains 'you'); where the sentences is uttered (if the sentence contains 'here') and so on. A natural thought was to build into indices features for all such expressions, so that indices would contain all the features that go into determining extensions of expression. Thus, indices would be n-tuples of a world, time, place, speaker, addressee and so on. Lewis (1970) is a good example of an "index semantics" with indices containing many features. However, a number of developments resulted in such approaches being abandoned or at least significantly modified.

Hans Kamp (1971) discovered that in a language with standard feature-of-index shifting tense operators and contextually sensitive expressions that are sensitive to that same feature, such as 'now,' one needs two temporal coordinates. The point can be illustrated using a sentence in which 'now' occurs embedded under e.g. a past tense operator (assume 'one week ago' is a past tense operator):

15. One week ago Sarah knew she would be in Dubrovnik now.

When this sentence is evaluated at an index, there must be a time in the index for 'one week ago' to shift. The embedded sentence ('Sarah knew she would be in Dubrovnik now') is then evaluated relative to an index whose time feature has been shifted back one week. But then if 'now' takes that time as its value, we predict that 15 means that one week ago Sarah knew she would be in Dubrovnik *then*. But the sentence doesn't mean that. So the index must contain a second time, in addition to the one shifted by 'one week ago,' that remains unshifted so that the embedded occurrence of 'now' can take it as its value.

Kamp's requirement of there being two time coordinates is sometimes called the requirement of *double indexing*. I emphasize again that the requirement stems from there being in the language an operator that shifts a certain feature (time, in our case) and a contextually sensitive expression that picks up as its value the same feature. The argument above given for double indexing of times, then, assumes that temporal expressions ('One week ago') are index shifting operators. Many, including the present author, doubt this claim. (See King [2003] for discussion.) But similar arguments (involving 'actual' and 'Necessarily') could be given for double indexing of worlds.

At any rate, on the basis of such considerations, it was thought that minimally, one needed two indices, each of which contained (at least) a world and a time. However it was Kaplan (1989) (written in the early 1970s and circulated for years in mimeograph form) that provided the proper theoretical understanding of double indexing. Kaplan forcefully argued that not only do we need two indices for the reasons Kamp suggested as well as others (see section VII of 'Demonstratives'), but we need to recognize that the indices are representing two very different things, with the result that we need to recognize two different kinds of semantic values. One index represents *context of utterance*. This is the index that provides values for contextually sensitive expressions such as 'I,' 'now,' 'here' and so on. The intuitive picture is that a sentence taken relative to a context of utterance has values assigned

to such contextually sensitive expressions. This results in the sentence having a *content*, *what is said by the sentence*, taken in that context.

So If I utter 'I am hungry now' on June 12, 2006, the content of the sentence in that context, what I said in uttering it then, is that Jeff King is hungry on June 12, 2006. Now that very content can be evaluated at different *circumstances of evaluation*, which are what the other index represents. For simplicity, think of circumstances of evaluation as simply possible worlds. Then we can take the sentence 'I am hungry now' and consider its content relative to the context of utterance described above. That content, or proposition, can then be evaluated for truth or falsity at different circumstances of evaluation (possible worlds). It is true at worlds in which Jeff is hungry on June 12, 2006 and false at those where he is not.

This distinction between context and circumstance, which the two indices represent, gives rise to a distinction between two kinds of semantic value (here we confine ourselves to the semantic values associated with sentences). On the one hand, the sentence 'I am hungry now' has a meaning that is common to utterances of it regardless of speaker or time. It is this meaning that determines what the content of that sentence is taken relative to contexts with different speakers and times. So this meaning, which Kaplan called *character*, determines a function from contexts to propositional content or what is said. By contrast, there is a sense in which the sentence 'I am hungry now' uttered by me now and Rebecca tomorrow means different things. This is because the sentence has different *contents* relative to those two contexts. So content is the other kind of semantic value had by sentences. Contents are true or false at worlds, so contents determine functions from worlds to truth values. In summary, character determines a function from context to content; content determines a function from worlds to truth values. Kaplan's distinction between context and circumstance and the corresponding distinction between character and content has been hugely influential and widely accepted.

Another important feature of Kaplan's (1989) work is his argument that both *demonstratives* (contextually sensitive words whose use requires the speaker to do something like demonstrate (point at) who she is talking about: 'he,' 'she,' 'this,' 'that') and *pure indexicals* (contextually sensitive words that don't require such demonstrations: 'I,' 'today,' etc.) are *devices of direct reference*. If we think of contents of sentences, propositions, as structured entities having as constituents the individuals, properties and relations that are the contents (relative to a context)

of the expressions in the sentence, a view Kaplan likes, we can understand the claim that indexicals and demonstratives directly refer as the claim that these expressions contribute to propositions (relative to a context) the individuals they refer to (in the context). Thus, when I say: 'I am hungry,' the indexical 'I' contributes me to the proposition expressed by that sentence in that context.

Historically, the importance of this direct reference account of indexicals and demonstratives is its anti-Fregean thrust. Recall that for Frege, expressions generally, even those that refer to individuals, contribute to propositions senses that pick out their references and not the references themselves. In claiming that indexicals and demonstratives contribute individuals to propositions rather than senses that pick out those individuals, Kaplan was proposing a radically anti-Fregean account of indexicals and demonstratives. Kaplan's arguments here complemented the anti-Fregean arguments of one of the most influential works in philosophy of language of the twentieth century: Saul Kripke's (1980) *Naming and Necessity*.

Among other things, Kripke (1980) provided powerful arguments against what he sometimes calls *the description theory* of names. On the description theory, names are held to be both synonymous with definite descriptions and (more weakly) to have their references fixed by definite descriptions. So, for example, 'Aristotle' might be thought to be synonymous with 'the teacher of Alexander,' and whoever satisfies this description is the referent of 'Aristotle.' Frege's view was thought to be a version of the description theory, since Frege seems to say that the sense of a proper name can be expressed by a definite description (Frege [1892a] note B), in which case the name and descriptions would be synonymous. Kripke argued very compellingly that descriptions were neither synonymous with, nor determined the reference of, proper names. As to synonymy, Kripke pointed out that whereas

16. The teacher of Alexander taught Alexander.

expressed (nearly) a necessary truth,

17. Aristotle taught Alexander.

expresses a highly contingent truth. But if the name and description were synonymous, the two sentences should be synonymous and so both should be contingent or both should be necessary. But they aren't. Indeed, the name and description seem to function very differently semantically. As Kripke famously noted, whether 17 is true at any possible world depends on the properties of Aristotle at that world. This because 'Aristotle' is what Kripke called a *rigid designator*: the expression designates Aristotle at

every world where he exists, and never designates any individual other than Aristotle. Hence evaluating the sentence at a world always requires us to check Aristotle's properties there. By contrast, 'the teacher of Alexander' presumably designates different individuals at different worlds, depending on who taught Alexander there. Thus, this expression is *non-rigid*.

As to descriptions determining the referents of names, Kripke adduced a number of considerations but perhaps the most persuasive was the following. Consider a name and any description that allegedly fixes the referent of the name, say 'the man who proved the completeness of arithmetic' fixes the referent of 'Godel.' If we imagine that in fact some man Schmidt satisfies the description, we do not conclude that he is the referent of 'Godel.' Quite the contrary, we conclude that the referent of 'Godel,' that is, Godel, fails to satisfy the description. But then the description does not fix the referent of the name (i.e. the referent is not whoever satisfies the description).

The arguments of Kaplan (1989) and Kripke (1980), together with arguments given by Donnellan, Marcus, Putnam and others turned semantics in a very anti-Fregean direction from the 1970s on. This anti-Fregean strain as applied to singular terms is sometimes called *the new theory of reference*.

As we saw above, Kaplan claimed that indexicals and demonstratives were directly referential and contributed their referents (relative to a context) to the propositions expressed by sentences in which they occur (interestingly, this is not reflected in Kaplan's [1989] formal system, which makes use of unstructured propositions that have no constituents corresponding to the words in the sentences that express the propositions; but his informal remarks make clear his intent). By contrast, though Kripke (1980) argued against the descriptive theory of names, he cautiously made no positive claims about what names contribute to propositions (the preface to Kripke [1980] makes clear that this caution was intended—see pp. 20–21). In a series of works in the 1980s, most famously Salmon (1986) and Soames (1987), Scott Soames and Nathan Salmon offered powerful arguments in favor of the view that names too were devices of direct reference and contributed only their bearers to propositions expressed by sentences in which they occur. Both Soames and Salmon defended the view that sentences (relative to contexts) express structured propositions, with names (and indexicals and demonstratives) contributing the individuals to which they refer to proposi-

tions. Salmon and Soames both also thought that attitude ascriptions such as the following:

18. Nathan believes that Mark Twain is an author.

assert that the subject (Nathan) stands in a certain relation (expressed by 'believes') to a structured proposition (expressed by the embedded sentence). If that is right and if names contribute only individuals to propositions expressed by sentences in which they occur, then (assuming a simple principle of compositionality) 18 expresses the same proposition as

19. Nathan believes that Sam Clemens is an author.

Thus, on the Soames-Salmon view 18 and 19 cannot differ in truth value. Though this seems counterintuitive, Soames (1987) and Salmon (1951) offer spirited defenses of this result. Soames (1987) also offers extremely compelling arguments against the view that propositions are unstructured sets of worlds (or circumstances). Some version of the Soames/Salmon view is widely considered to be the standard direct reference view in semantics. Views such as theirs, which make use of structured propositions and endorse direct reference for names, demonstratives and indexicals, are often called *Russellian*.

About the same time the new theory of reference was becoming prominent, quite different developments were taking place in semantics. In pioneering work first presented in the late 1960s (as the William James Lectures at Harvard; later published in Grice [1989] as Essay 2), Paul Grice sought to give a (somewhat) systematic account of (as we would now put it) how the production of a sentence with a certain semantic content can convey further information beyond its semantic content. To give an example from Grice, suppose A and B are planning their itinerary for a trip to France and both know A wants to visit C if doing so wouldn't take them too far out of their way. They have the following exchange:

A: Where does C live?

B: Somewhere in the south of France.

Since both are aware that B offered less information than is required for the purposes at hand, and since B can be presumed to be attempting to cooperate with A, B conveys that she doesn't know where C lives, though this is no part of the semantic content of the sentence she uttered. Grice gave an account of how such information (not part of the semantic content of any sentence asserted) can be conveyed. The account depended on the claim that conversational participants are all obeying certain principles in engaging in conversation. The main idea, as illustrated above, is that conversational partici-

pants are trying in some way to be cooperative, and so to contribute to the conversation at a given point what is required given the purpose and direction of the conversation. Grice's central theoretical idea was that certain types of information exchange and certain types of regularities in conversations don't have purely semantic explanations. The study of how information gets conveyed that goes beyond the semantic content of the sentences uttered falls in the field of *pragmatics*, (which is why, though Grice's work is extremely important, it hasn't been discussed more in an entry on semantics).

In a series of papers that (for our purposes anyway) culminated in Stalnaker (1978), Robert Stalnaker, consciously following Grice, was concerned with ways in which in conversations information can be conveyed that goes beyond the semantic contents of sentences uttered as a result of conversational participants obeying certain principles governing conversation. More specifically, Stalnaker developed an account of how context of utterance and semantic contents of sentences (relative to those contexts) produced in those contexts can mutually influence each other.

Of course, how context influences the semantic content of sentences relative to those contexts was already fairly well understood. As discussed above, for example, context supplies the semantic values relative to those contexts for contextually sensitive expressions such as 'I.' Stalnaker sought to understand how the content relative to a context of a sentence uttered can affect the context. Stalnaker began by introducing the notion of *speaker presupposition*. Stalnaker understood the proposition expressed by a sentence (relative to a context) to be a set of possible worlds (the set of worlds in which the sentence taken in that context is true). Very roughly, the propositions a speaker presupposes in a conversation are those whose truth he takes for granted and whose truth he thinks the other participants take for granted too.

Consider now the set of possible worlds that are compatible with the speaker's presuppositions (the set of worlds in which every presupposed proposition is true). Stalnaker calls this *the context set*, and it is for him a central feature of a context in which a conversation occurs. (Strictly, every participant in the conversation has his own context set, but we will assume that these are all the same—Stalnaker calls this a *non-defective context*.) They contents of sentences (relative to a context) affect the context in the following way: if a sentence is asserted and accepted, then any world in the context set in which the sentence (taken in that context) is false is eliminated from the context set. In short, (accepted) assertions function to

reduce the size of the context set, or eliminate live options.

Stalnaker uses this idea to explain a variety of phenomena, including how the utterance of sentences with trivial semantic content (relative to a context) can nonetheless be informative. It is important to see that Stalnaker, like Grice, took his account here to be not part of semantics, but rather to be something that presupposed the semantics of sentences (taken in contexts). In short, like Grice's work, it was work in pragmatics. However, Stalnaker's idea that the information conveyed by the utterance of multiple sentences in a discourse can go beyond anything countenanced by traditional semantics and that it is important to understand the dynamics of conversation to understand how information is conveyed influenced others who went on to develop *semantic* theories that capture the dynamics of conversation, (Lewis [1979] was another important early influence to the same effect).

In the early 1980s, Irene Heim (1982) and Hans Kamp (1981) independently arrived at very similar semantic accounts that were intended to apply to multi-sentence discourses. Kamp's view is called *Discourse Representation Theory (DRT)* and Heim's view is sometimes called that or *File Change Semantics (FCS)*. To take a simple example of the sort that DRT and FCS were designed to handle, consider a (short) discourse such as:

20. Alan owns a donkey. He beats it.

Using Kamp's formulation, the *discourse representation structure (DRS)* associated with the first sentence of 20 would (roughly) look as follows:

$x_1 x_2$   
 $x_1 = \text{Alan}$   
 $\text{donkey}(x_2)$   
 $x_1 \text{ owns } x_2$

After the utterance of the second sentence, the DRS associated with the entire discourse would look as follows, where we have simply added one more line (*a condition*) to the DRS for the first sentence of 20 (we assume that 'He' is anaphoric on 'Alan' and 'it' on 'a donkey'):

$x_1 x_2$   
 $x_1 = \text{Alan}$   
 $\text{donkey}(x_2)$   
 $x_1 \text{ owns } x_2$   
 $x_1 \text{ beats } x_2$



Note that expressions like ‘a donkey’ introduce variables (called *discourse referents*) and predicates (‘donkey’) into DRS’s and not existential quantifiers. Again very roughly, this DRS (and hence the original discourse) is true in a model iff there is an assignment to the variables of the DRS that results in all its conditions being true in that model. It is the requirement that *there is* such an assignment that results in default existential quantification of free variables. So though indefinites like ‘a donkey’ are not existential quantifiers on this view, they have existential force (in this case, anyway) due to default existential quantification of free variables. Aside from the desire to apply semantics at the level of discourse instead of sentence, much of the motivation for DRT and FCS came from cases such as 20 (and others) in which a pronoun is anaphoric on another expression (see entry on *anaphora*).

DRT and FCS led directly to the development of other semantic accounts designed to capture the dynamics of conversation. In the paper that initiated what is now often called *dynamic semantics*, Groenendijk and Stokhof (1991) make clear that they see their account as a descendant of DRT and throughout the paper they compare their *Dynamic Logic (DL)* account to DRT. The basic idea of DL is that instead of thinking of expressions as having “static” meanings, think of meanings as things that given inputs, produce outputs. A bit more formally, think of the meanings (in models) of formulae of first order logic as given by the sets of assignments to variables that satisfy the formulae. So for example, the meaning of ‘Fx’ in a model M is the set of all assignments such that they assign to ‘x’ something in the extension of ‘F’ in the model M. Dynamic logic claims that the meaning of a formula in first order logic is a set of *pairs* of assignments: the first, the input assignment; the second, the output assignment. For “externally dynamic” expressions (e.g. conjunction, existential quantifiers), these can differ and the result is that interpreting these expression can affect how subsequent expressions get interpreted. For since the output assignments can be different from the input assignments for these dynamic expressions, and since the output of these expressions may be the input to subsequent expressions, the interpretation of those subsequent expressions may be affected.

There is currently much research being done within the framework of dynamic semantics, particularly among linguists. Muskens, van Benthem and Visser (1997) provide a good general overview.

There are many important topics in semantics that could not be covered in the present article. These include

the theory of generalized quantifiers, the semantics of conditionals, the semantics of non-declarative sentences, the semantics of metaphor and two dimensional semantics. Interested readers are encouraged to pursue these matters on their own.

**See also** Carnap, Rudolf; Conversational Implicature; Davidson, Donald; Frege, Gottlob; Grice, Herbert Paul; Heidegger, Martin; Hempel, Carl Gustav; Hintikka, Jaakko; Kaplan, David; Kripke, Saul; Lewis, Clarence Irving; Lewis, David; Logical Positivism; Marcus, Ruth Barcan; Meaning; Modality, Philosophy and Metaphysics of; Montague, Richard; Pragmatics; Putnam, Hilary; Reference; Russell, Bertrand Arthur William; Syntax; Tarski, Alfred; Wittgenstein, Ludwig Josef Johann.

### **Bibliography**

- Ayer, A.J., ed. *Logical Positivism*. Glencoe, IL: Free Press, 1959.
- Beaney, Michael, ed. *The Frege Reader*. Oxford U.K.; Cambridge, MA: Blackwell, 1997.
- Bennett, Michael. “Some Extensions of a Montague Fragment.” PhD diss, UCLA, 1974.
- Carnap, Rudolf. “The Elimination of Metaphysics through The Logical Analysis of Language” (1932). In *Logical Positivism*, edited by A.J. Ayer. Glencoe, IL: Free Press, 1959.
- Carnap, Rudolf. *Meaning and Necessity, A Study in Systematics and Modal Logic*. Chicago: University of Chicago Press, 1947.
- Church, Alonzo. “A Formulation of the Logic of Sense and Denotation.” In *Structure, Method, and Meaning: Essays in Honor of Henry M. Sheffer*. New York: Liberal Arts Press, 1951.
- Church, Alonzo. “On Carnap’s Analysis of Statements of Assertion and Belief.” *Analysis* 10 (1950): 97–99.
- Davidson, Donald. “On Saying That.” *Synthese* 19 (1968): 130–146
- Davidson, Donald. “Truth and Meaning.” *Synthese* 17 (1967): 304–323.
- Frege, Gottlob. “Function and Concept” (1891). In *The Frege Reader*. Oxford U.K.; Cambridge, MA: Blackwell, 1997.
- Frege, Gottlob. “On Concept and Object” (1892). In *The Frege Reader*. Oxford U.K.; Cambridge, MA: Blackwell, 1997.
- Frege, Gottlob. “On Sense and Reference” (1892). In *The Frege Reader*. Oxford, U.K.; Cambridge, MA: Blackwell, 1997.
- Grice, Paul. *Studies in the Ways of Words*. Cambridge, MA: Harvard University Press, 1989.
- Groenendijk, J., and M. Stokhof. “Dynamic Predicate Logic.” *Linguistics and Philosophy* 14 (1991): 39–100.
- Heim, Irene. “The Semantics of Definite and Indefinite Noun Phrases.” Doctoral thesis, University of Massachusetts, 1982.
- Hempel, Carl. “The Empiricist Criterion of Meaning” (1950). In *Logical Positivism*. Glencoe, IL: Free Press, 1959.
- Hintikka, Jaakko. “Modality and Quantification.” *Theoria* 27 (1961): 110–128.

- Hodges, Wilfrid. "Tarski's Truth Definitions." *The Stanford Encyclopedia of Philosophy*, winter 2001 ed. Available from <http://plato.stanford.edu/archives/win2001/entries/tarski-truth/>.
- Kamp, Hans. "Formal Properties of 'Now.'" *Theoria* 37 (1971): 227–273.
- Kamp, Hans. "A Theory of Truth and Semantic Representation." In *Formal Methods in the Study of Language*, edited by J. Groenendijk and M. Stokhof. Amsterdam: Mathematical Centre, 1981.
- Kanger, Stig. *Provability in Logic*. Stockholm: Almqvist & Wilksell, 1957.
- Kaplan, David. "Demonstratives." In *Themes from Kaplan*, edited by Joseph Almog, John Perry, and Howard Wettstein. New York: Oxford University Press, 1989.
- King, Jeffrey C. "Tense, Modality, and Semantic Values." *Philosophical Perspectives* 17 (2003): 195–245.
- Kripke, Saul. "A Completeness Theorem in Modal Logic." *The Journal of Symbolic Logic* 24 (1) (1959): 1–14.
- Kripke, Saul. "Semantical Considerations on Modal Logic" (1963). In *Reference and Modality*, edited by Leonard Linsky. London: Oxford University Press, 1971.
- Kripke, Saul. *Naming and Necessity* (1972). Cambridge, MA: Harvard University Press, 1980.
- Lewis, C. I., and C. H. Langford. *Symbolic Logic*. New York: Century, 1932.
- Lewis, David. "General Semantics." *Synthese* 22 (1970): 18–67.
- Lewis, David. "Scorekeeping in a Language Game." *Journal of Philosophical Logic* 8 (1979): 339–359.
- Montague, Richard. "Logical Necessity, Physical Necessity, Ethics, and Quantifiers" (1960a). In *Formal Philosophy; Selected Papers of Richard Montague*, edited by Richmond Thomason. New Haven, CT: Yale University Press, 1974.
- Montague, Richard. "On the Nature of Certain Philosophical Entities" (1960b). In *Formal Philosophy; Selected Papers of Richard Montague*, edited by Richmond Thomason. New Haven CT: Yale University Press, 1974.
- Montague, Richard. "The Proper Treatment of Quantification in Ordinary English" (1973). In *Formal Philosophy; Selected Papers of Richard Montague*, edited by Richmond Thomason. New Haven, CT: Yale University Press, 1974.
- Muskens, Reinhard, Johan van Venthem, and Albert Visser. "Dynamics." In *Handbook of Logic and Language*, edited by Johan van Benthem and Alice ter Meulen. Cambridge, MA: MIT Press, 1997.
- Russell, Bertrand. *Principles of Mathematics*. 2nd ed. Cambridge, U.K.: Cambridge University Press, 1903.
- Salmon, Nathan. *Frege's Puzzle*. Cambridge, MA: MIT Press, 1986.
- Soames, Scott. "Direct Reference, Propositional Attitudes, and Semantic Content." *Philosophical Topics* 15 (1987): 47–87.
- Stalnaker, Robert. "Assertion" (1978). In *Context and Content; Essays on Intentionality in Speech and Thought*. New York: Oxford University Press, 1999.
- Tarski, Alfred. "Der Wahrheitsbegriff in den formalisierten Sprachen." *Studia Philosophica* 1 (1935): 261–405.
- Weinstein, Scott. "Truth and Demonstratives." *Nous* 8 (1974): 179–184.

Jeffrey C. King (2005)

## SEMANTICS, HISTORY OF

*The scope of this article is in part determined by the following restrictions. (1) Although the development of semantics in the twentieth century equals or surpasses all that was done earlier, it receives very little attention here because the major theories and theorists of this period are thoroughly discussed in other articles. (2) The only semantic theories considered are those developed by Western philosophers; thus, no account is taken of the theories of meaning propounded, for example, by ancient Hindu philosophers or by European grammarians or linguists. (3) Since semantic theories concerning nonlinguistic signs tend to involve considerations of theories of knowledge generally, they are not discussed here except as they may occasionally bear directly on a theory of linguistic meaning. On the other hand, much of what philosophers have had to say about language is discussed here, whether or not it can be precisely described as semantics.*

The contents of this article are arranged as follows.

### ANTIQUITY

THE COSMOLOGISTS  
THE SOPHISTS  
CONVENTIONALISM AND NATURALISM  
PLATO  
ARISTOTLE  
THE STOICS  
THE EPICUREANS

### THE MIDDLE AGES

ST. AUGUSTINE  
BOETHIUS  
ST. ANSELM  
ABELARD  
IMPOSITIONS AND INTENTIONS  
"SCIENTIA SERMOCINALIS"  
THE PROPERTIES OF TERMS  
SYNCATEGOREMATA  
SPECULATIVE GRAMMAR

### THE RENAISSANCE AND ENLIGHTENMENT SEMANTICS, LOGIC, AND EPISTEMOLOGY

BACON  
HOBBS  
THE "PORT-ROYAL LOGIC"  
LOCKE  
LEIBNIZ  
BERKELEY  
MAUPERTUIS AND HIS CRITICS  
CONDILLAC  
LAMBERT, HAMANN, AND HERDER  
THE "IDÉOLOGUES"  
UNIVERSAL GRAMMAR

## THE NINETEENTH AND TWENTIETH CENTURIES

BENTHAM  
HUMBOLDT  
JOHNSON  
MILL  
PEIRCE AND THE PRAGMATISTS  
FREGE  
MAUTHNER  
HUSSLER AND MEINONG

## ANTIQUITY

### THE COSMOLOGISTS

Since the earliest Greek philosophers were primarily cosmologists, their views on language are not the most fully developed (or best preserved) of their doctrines. Sources very late in antiquity attributed to Pythagoras (fl. 530 BCE) the view that although the soul assigned names to things, it did so not arbitrarily but on the basis of a natural connection between them, somehow like that between mental images and their originals. Modern historians sometimes credit Heraclitus (fl. 500 BCE) with having thought a great deal about language, but most of the fragments offered in evidence have to do with the *logos*, which surely is to be interpreted as the guiding principle of nature rather than as word or language. While we have nothing of his explicitly on language, it seems likely that Heraclitus did attach philosophical significance to the puns or contradictions in terms on which some of his paradoxical remarks depend.

Semantic theory seems to have made its first definite appearance in philosophy in the monism of Parmenides (fl. 475 BCE), who maintained that only what was true was expressible. He evidently based this remarkable doctrine on the argument that a statement is false if and only if it contains a false name, but a false name is by definition a name lacking a real bearer and hence a name that names or expresses nothing. (His monism of course entailed that there was only one real name-bearer.) Thus he described several words, such as “becoming” and “perishing,” as “mere names that mortals have established, believing them to be true”—that is, believing that there really are such processes, which Parmenides denied.

### THE SOPHISTS

Language first became a subject of specialized inquiry among the Sophists, who, unlike their philosophical predecessors, were more interested in man than in the cosmos. That orientation alone would probably have drawn them to the study of language, but there was also

the fact that they earned their livings teaching people to speak well. Economic as well as philosophical considerations therefore probably played a part in leading them to include at least grammar as an important part of their work. Protagoras (fl. 445 BCE), the first of the Sophists, may also be considered the first grammarian. He distinguished the tenses and something like grammatical moods (classifying sentences as answers, questions, commands, and wishes), and he classified nouns as masculine, feminine, and “inanimate” (a division based on semantical rather than syntactic considerations, since it depended on the particular sex or lack of sex in the things the nouns were used to name). Grammar developed rapidly among the Sophists. Among the more philosophically interesting parts of grammatical theory to be found in Plato, who doubtless learned much of it from the Sophists, are distinctions between subject and predicate, between substantive and adjective, between the active and passive voices, and among types of discourse—political, rhetorical, conversational, dialectical, and technical.

The Sophists originated semantical as well as grammatical inquiries. Prodicus (fl. 435 BCE), who Plato thought was the best of the Sophists on language, seems to have operated on the hypothesis that there were no genuine synonyms, that where there were two words, there were two meanings. In Plato’s dialogues Prodicus is depicted drawing instructive distinctions between “enjoyment” and “pleasure,” “esteem” and “praise,” “fearlessness” and “courage,” for example; and he insisted on the study of “the right use of words” as the beginning of education. Protagoras, Prodicus, and Hippias (fl. 435 BCE) are all credited with treatises on “the correctness of names,” and Socrates (d. 399 BCE) is depicted discoursing on that subject in Xenophon’s *Memorabilia* (III, xiv, 2–7).

Semantics may have become a theoretical issue for the first time in the paradoxical arguments propounded by Gorgias (fl. 435 BCE) in support of his third nihilistic thesis. The three theses were (1) nothing exists; (2) even if something existed, it would be unknowable; (3) even if something existed and were knowable, it would be incommunicable. Gorgias gave four arguments for thesis (3) along the following lines. Suppose there really are things and they can be perceived by our senses. Then (a) some of those things will be perceivable by one sense only and others by another sense only; and since one sense cannot perceive objects proper to another sense, a system of audible signs will not permit communication regarding things perceivable only by sight, and so on for the other senses. In any case, (b) those supposed things are not identical with any signs one might use to communi-

cate about them, and so one could never convey the things themselves to another person but only the signs.

Moreover, (c) even if one could produce signs exactly representing those supposed things, he could not communicate those signs to another person, for the signs themselves are things, and no one can have in his mind the same things that someone else has in *his* mind at the same time. Finally, (d) any signs we might use would have to be formed as a result of our perception of those supposed things, but since genuine knowledge of a cause cannot be gained from its effect, no knowledge of those things could be communicated by means of any signs. Occasionally in arguments (a), (b), and (c) Gorgias seems, like Jonathan Swift's Laputans, to have sophistically confused talking about things with handing them around; but not all his paradoxes of communication are transparent, and some passages in Plato and Aristotle suggest that Gorgias's arguments may have helped to shape their semantic theories.

### CONVENTIONALISM AND NATURALISM

The oldest surviving arguments in support of a particular semantic theory may be those attributed very late in antiquity to Democritus (fl. 420 BCE), perhaps presented originally in his book *On Words*. He is supposed to have offered the following four considerations in support of his position that the relation between names and things named is conventional (*θέσει*) rather than natural (*φύσει*): (a) the occurrence of homonyms, that is, one and the same name for things different in nature; (b) the occurrence of synonyms, that is, different names for one and the same thing; (c) the occurrence of name-changes while the thing named remains the same in nature; (d) the nonoccurrence of verbal analogies corresponding with real analogies, for instance, there is a verb analogous to the noun "understanding" but none analogous to "justice."

In all probability no philosopher ever held a thoroughgoing semantic naturalism, although there are traces of tendencies in that direction in the doctrine attributed to Pythagoras and in the assumptions that appear to underlie the work of Prodicus and Gorgias. The opposition of naturalism and conventionalism as semantic theories forms the point of departure for the development of Plato's semantics of names in the *Cratylus*. Much of the significance of the *Cratylus* and of ancient philosophy of language generally has been obscured, from antiquity onward, by the confusion of this semantic issue with a dispute over the origin of language in which "naturalism" and "conventionalism" were the principal doctrines. In

that dispute, however, it was not the naturalist but the conventionalist position that was preposterous, conventionalism in that context being the claim that language first arose as a result of agreements among men or because some especially powerful individual compelled those around him to use his names for things. There are, of course, implications for semantics in theories about the origin of language, but neither Plato nor any other ancient philosopher of the first rank failed to distinguish between the two inquiries.

### PLATO

The oldest surviving work of any kind on language is Plato's *Cratylus* (probably written about 388 BCE). The main topic of this dialogue is the nature of the relation between names and things named.

**THE CRATYLUS.** At the beginning of the *Cratylus* a kind of semantic naturalism is attributed to Cratylus and a kind of semantic conventionalism to Hermogenes. All that is said about naturalism at the outset is that it seems unintelligible, and the first serious undertaking is a discussion of the conventionalism advanced by Hermogenes in these words:

I cannot be persuaded that there is any correctness of name other than convention [*ξυνθήκη*] and agreement [*δμολογία*]. For it seems to me that whatever name anyone gives to a thing is the correct one, and if someone changes that name for another, the later one is no less correct than the earlier—just as when we change the names of our slaves. For no name has been generated by nature for any particular thing, but rather by the custom [*νόμος*] and usage [*ἔθει*] of those who use the name and call things by it. (384 C–D)

There is nothing in this conventionalism we have not already seen in the Democritean arguments except the claim that "whatever name anyone gives to a thing is the correct one," and Socrates immediately asks whether this claim is intended to apply to private persons as well as to nations (385A). Hermogenes fails to appreciate the difference, and when, as a result, Socrates is on the point of showing that this subjectivist claim destroys the possibility of distinguishing between true and false statements, Hermogenes tries to salvage it by suggesting an analogy between arbitrary individual name-giving and different natural languages (385D–E). The picture presented is that of a conventionalist who recognized that the existence of different autonomous natural languages was strong con-

firmation of his position and was then so carried away as to produce a doctrine of autonomous idiolects, evidently reasoning as follows: Just as the Greek word for horse is no more and no less correct than the Persian word for horse, so there is no basis for correcting a Greek who should decide to use the Greek word *anthropos* where other Greeks use *hippos* and vice versa.

The conventionalism presented as a basis for discussion in the *Cratylus* is entirely plausible except for the obviously untenable doctrine of autonomous idiolects. One consequence of the doctrine is that at any given time a given thing (or type of thing) has just as many correct names as there are people who name it differently (385D). This suggests some sort of Protagorean skepticism in its author, but Hermogenes is ready to agree that “things have some fixed reality of their own, not relative to us or caused by us” (386D). Socrates uses this admission to show the necessity of objectively correct names. There are real things, he says, and real things are not subject to our whims. We recognize that we cannot do certain jobs involving real things simply by fiat. We must make the correct moves, using the correct tools, and the correct tools for a given job cannot be generally described as the first ones anyone may choose (386E–387B). Now in the use of language, names are our tools, and we employ those tools in doing two essential jobs plainly involving real things: “teaching” (communicating the truth) and “classifying things according to their natures” (387B–388B). If “whatever name anyone gives to a thing is the correct one,” we clearly have no chance of succeeding in communicating the truth to one another or in developing classification schemes that will “carve reality at the joints.”

The destruction of the doctrine of autonomous idiolects leaves a gap in conventionalism, a gap that was there in any case but that would not have been so easily seen if Plato had not thus deliberately marred this conventionalism in order to call attention to it. Not just anyone is an arbiter of the correctness of names; but then “who *does* provide us with the names we use?” (388D). The answer is derived from the sounder portions of Hermogenes’s conventionalism, in which he claimed that custom or law generates our names for things. This suggests that the arbiter of custom, or the lawgiver (*νομοθέτης*), may be identified as the name-maker (*ὀνοματουργός*) (388D–389A). The “law-giver” is Plato’s personification of a recognized stipulative linguistic authority, more nearly like the French Academy or the *Oxford English Dictionary* than like an individual—Solon, for instance.

This refurbished conventionalism is adequate as far as questions of pronunciation, word order, and usage are concerned; these can be settled by having recourse to the recognized authority. The question raised by the criticism of autonomous idiolects, however, was not, “how do we determine which names are *accepted*?” but, rather, “how do we determine which names are *correct*?” Plato took the two questions to be distinct and made his most important contribution to the semantics of names in answering the second of them. The development of his answer may be traced out as follows.

If the refurbished conventionalism is to do any more than offer an account of the phenomena of a language, it must be augmented by part of Cratylus’s naturalism, which was originally stated in these three claims: “(a) for each of the things that really exist there is a correctness of name that has been produced by nature; (b) that is not a name which some people agreeing together to give as a name do give as a name, uttering a bit of their voice in accordance therewith; but (c) there is a kind of correctness of names that is the same for all, both Greeks and barbarians” (383A–B).

At the beginning of the dialogue this position was taken to be unintelligible because it was thought to be in competition with conventionalism as an account of the phenomena of a language. Claim (b) does seem to justify the view that the theory is just a wrongheaded account of that kind. Temporarily ignoring claim (b), Socrates proceeds to show that this naturalism makes sense as an account of the conceptual underpinnings of all languages. The fact that the word for horse in Greek is “hippos” and in Latin “equus” shows that different linguistic authorities are operative in different natural languages. Both those words are perfectly acceptable, intertranslatable names for horse; and what makes them so is the fact that each of them embodies in different marks (or sounds) a single “ideal name,” which belongs to horse “by nature,” whose correctness has been produced by nature, and which is the same for all, both Greeks, who say “hippos,” and “barbarians,” who say “equus” (389C–390A). That single ideal name cannot be the type of which occurrences of “hippos” or of “equus” are the tokens, since it is “the same for all.” Nor can it be identified with what Plato called the form of horse, for although the form of horse may be the ideal *horse*, there is nothing of which it could conceivably be a *name*. Instead, the ideal name embodied as well in “equus” as in “hippos” is the correctly framed concept *horse*, and the difference between the two words is merely the difference between two equally good notations. To say that the concept is

framed correctly is to say that it is the concept of the form rather than of individuals participating in that form; to say that its correctness has been produced by nature is to say that it somehow resembles the form. The correctly framed concept *horse* is a logically proper name of the form of horse; it is the ideal name for which all the words correctly translatable into English as “horse” are various notations.

Plato goes on to develop and apply this theory along the following lines. If we should come across a natural language the speakers of which owned horses and cows but had only one name for both species, or had no single name for horse, using instead an indifferently ordered string of names for legs, head, tail, and so on, then we should have a genuine case of incorrect names. The speakers of that language would be laboring under the influence of incorrectly framed concepts, concepts that fail to carve reality at the joints. Thus, we avoid incorrect names (such as “phlogiston”) to the extent to which philosophy and science (personified by Plato as “the dialectician”) have provided us with a correct conceptual schema (390C–E).

But the embodiment even of correctly framed concepts in the evolving phenomena of a natural language will sooner or later lead to the development of homonyms and synonyms, which, although not incorrect, are infelicitous for the purposes of science and philosophy. Such infelicities could be avoided if we were to construct a precise, consciously designed concept-notation for the use of philosophers and scientists (421E–423E, 424D–425A). And even if we do not or cannot actually construct it, the notion of a perfectly systematic embodiment of correctly framed concepts may serve as an ideal against which to measure the adequacy of technical language (435C). Thus, the frequently recurring project of an ideal language is to be found for the first time in the very first extant treatise on language.

Perhaps the single most unusual feature of this remarkable semantic theory is the doctrine of the ideal name. Within the *Cratylus* itself the identification of the ideal name with the correctly framed concept is not explicit, although it is clearly implied. That implication is strengthened by the many passages in other dialogues in which Plato did treat concepts as a kind of name—for instance, *Theaetetus* 189E, 206D; *Sophist* 263E; *Philebus* 38E–39A; *Phaedrus* 276A.

*Cratylus*’s naturalism and Hermogenes’ conventionalism are so expressed in the dialogue as to give every appearance of being simply Plato’s devices for raising semantic questions. Each of them contains an obvious,

completely gratuitous overextension. (Later in the dialogue [428A ff.] *Cratylus*’s claim [b] goes the way of Hermogenes’ autonomous idiolects.) Neither position alone is remotely plausible or likely to have been actually held by any philosopher, but each of them contains an essential ingredient of Plato’s own semantics of names.

**THE PARMENIDES AND THE SOPHIST.** Plato’s other major contributions to semantics occur in the later dialogues *Parmenides* and *Sophist*, in which he goes beyond the doctrine of the *Cratylus* in undertaking the connected tasks of (1) giving an account of the semantics of such names as lack existent bearers, (2) refuting the Parmenidean doctrine that false statements express nothing, and (3) giving an account of the semantics of simple statements.

(1) In *Parmenides* 160B–161A there is an attempt to state three necessary conditions for the meaningfulness of a denial of existence. (The example actually employed is the hypothesis “if a One does not exist,” which is eminently generalizable.) If we are meaningfully to say of  $x$  that it does not exist, then (a) “there is knowledge of”  $x$  (since “otherwise the very meaning of ... ‘ $x$ ’ does not exist” would be unknown”); (b)  $x$  is “something different from other things”; (c) “this non-existent [ $x$ ] has the characters of being *that*, and *something*, and of *being related to this*, or *to these*, and all other such characters.... If it does not exist, there is nothing against its having many characters; indeed it must [have many characters] if it is *this* [ $x$ ], and not another, that does not exist. If what is [said] not to exist is neither *the* [ $x$ ], nor *this*, and the statement is about *something else*, we ought not so much as to open our lips.” These three interdependent conditions do not seem inconsistent with the semantics in the *Cratylus*, and much of what was to be brought out later in the *Sophist* is already implicit in them—for instance, the distinction between existential and predicational occurrences of “is” (“if it does not exist, there is nothing against its having many characters”).

(2) When Parmenides or Plato speaks of expressing nothing, he means saying nothing meaningful, rather than saying nothing at all. This is implied in Parmenides’ fragments and is quite plain in Plato, when he says, for example, “Must we not assert that [a man] is not even expressing anything when he sets about uttering the words ‘a thing that is not?’” (*Sophist* 237E; *Cratylus* 429E). Those words constitute what Parmenides called a false name. A true, or meaningful, name is one having an identifiable existent bearer, a name that signifies something real; and there is no sharp semantical distinction between

true names and true statements—“If we are speaking the truth, evidently the things we are speaking of must be” (*Parmenides* 161E).

Thus the Parmenidean doctrine is that false statements are meaningless, or that truth and meaningfulness are indistinguishable. Although its scope was never restricted, the doctrine makes most sense when applied to statements of the form “*x* exists,” with which Parmenides was preoccupied. Such a statement, he would say, either is true or expresses nothing. In order to preserve the possibility of falsity, even in the limiting case of such statements, Plato had to question the Parmenidean dictum and establish that *what is not* has being in some respect (*Sophist* 241D), which he does in the complex, important doctrine of the interweaving of the Forms (252E–259C). However, his most direct answer to the Parmenidean doctrine is developed in his semantics of statements, an account based directly on the ontological theory just mentioned, since “any discourse we can have owes its existence to the weaving together of Forms” (260A).

(3) “Now, remembering what we said about Forms, ... let us consider words in the same way.... Words that when spoken in succession signify something, do fit together, while those that mean nothing when they are strung together do not” (261D–E). “Now a statement never consists solely of names spoken in succession, nor yet of verbs apart from names” (262A). Thus “the simplest and shortest possible kind” of statement is exemplified in “Theaetetus sits” or “Theaetetus flies,” “because ... it gives information about facts or events; ... it does not merely *name* something but *gets you somewhere* by weaving together verbs and names. Hence we say it *states* something” (262D). “Whenever there is a statement, it must be *about* something” (262E). Both the examples above are about one and the same existent thing, the bearer of the name “Theaetetus,” but the second is a combination of name and verb in which “what is different is stated as the same or [as is actually the case in this example] what is not as what is,” and anything “answering to that description finally seems to be really and truly a false statement” (263D).

In the *Parmenides* and *Sophist*, then, Plato not only extended semantics for the first time beyond the consideration of names to that of statements but, in doing so, also distinguished between meaningfulness and truth, showing for the first time that truth depends not merely on names but on certain syntactically regular combinations of verbs and names. It should be noted, however, that he does seem to have taken meaningfulness as the necessary and sufficient condition of grammaticalness.

Plato’s semantics of statements may be better appreciated against the background of the semantical doctrines of his contemporaries Antisthenes the Cynic (fl. 390 BCE) and Stilpo the Megarian (fl. 340 BCE). Beginning with the familiar “two names, two bearers” view, Antisthenes managed to reject all predication, on the grounds that what the subject named was one thing and what the predicate named was quite another, and to accept only identity statements of the form “*x* is *x*” or analogies of the form “*x* is like *y*.” Stilpo, too, rejected predication, perhaps on ontological grounds, since he insisted on “the unity of being” and may have thought that this could be expressed only in strict identity statements.

### ARISTOTLE

Aristotle’s primary interest in language was naturally that of a logician, and while his writings contain many passages on semantic questions, there is relatively little developed theory. His semantics of words (he treats of more than just names) is like Plato’s in many respects and is to be found mainly in *De Interpretatione*, Chapters 1–3. There he presents, with little or no argument, the following account of signification.

Although there are different natural languages, the people who use them are confronted with the same extramental things. The mental modifications arising from that confrontation are likenesses (*ὁμοιώματα*) of the things, and they are thus the same for all men too. Within a given natural language, written words are conventional symbols (*σύμβολα*) of spoken words. (Aristotle was no doubt unaware of ideographic notations.) The spoken words are, in turn, related to the mental modifications, first of all as symptoms, or natural signs (*σημεῖα*), of them—that is, of the presence of mental modifications in the speaker. More important, the spoken words are related to the mental modifications in the same way that written words are related to spoken words, as symbols of them. Just as written words constitute a conventional notation for (or embodiment of) spoken words, so do spoken words for mental modifications. Discussions of these passages have almost invariably failed to recognize the first of the two relations between spoken words and mental modifications as distinct and have confused the second relation with that of name to bearer.

It seems that, according to this account, words signify things in virtue of serving as symbols of mental modifications resembling those things. What sorts of “things” can words thus be made to signify? Not much is said on that question in *De Interpretatione*, but in *Categories* (Ch. 5) and *Sophistical Refutations* (Ch. 22), for example, vari-

ous words are said to signify (*σημαίνειν*) “a certain *this*,” “a qualification,” “a substance of a certain qualification,” “passivity,” “a certain relation to something else,” “a quantity,” and so on. More important, “‘man’ and every common name signifies not a certain *this*, but a quality or a relation or a mode (or something of the sort)” (*Sophistical Refutations* 178b38).

Ambiguity, Aristotle maintained, is theoretically unavoidable, for since “names and the sum-total of formulas [*λόγοι*] are finite while things are infinite in number ... the same formula and a single name must necessarily signify a number of things.” This will, however, give us no trouble unless “we think that what happens in the case of the names happens also in the case of the things, as people who are counting think of their counters,” which are in a one-to-one correspondence with the things counted (*Sophistical Refutations* 165a5). Although this passage is part of a warning against sophisms of ambiguity, when taken together with the preceding passage it seems to constitute an injunction against seeking the bearer of a common name, as Plato and so many of Aristotle’s successors did. A single individual is the bearer of many names in that they are all correctly predicable of it, but “we do not identify having one meaning with being predicable of one thing, since on that assumption even ‘musical’ and ‘white’ and ‘man’ [all of which are predicable of Socrates] would have one [and the same] meaning” (*Metaphysics* 1006b15).

The principal kinds of words recognized by Aristotle were the name (*ὄνομα*) and the verb (*ῥῆμα*—“predicate” is possibly a more accurate translation). He described them both as the smallest conventionally significant units, incapable of being true or false independently. A name without a bearer, such as “unicorn,” is neither “false” (as some of his predecessors had claimed) nor nonsignificant; and a name combined with “is,” “was,” or “will be” always produces something true or false. A verb uttered by itself is a name, but it additionally signifies time and “some combination, which cannot be thought of without the components.” Because of the latter additional signification, a verb “is always a sign (*σημεῖον*) of things being said of something else” (*De Interpretatione* 16b24, 16b7).

“Non-man” names nothing definite and so is not strictly a name; analogously, “does not walk” holds indifferently of all sorts of existents and nonexistents. These negated words Aristotle put into the separate categories of “indefinite names” and “indefinite verbs.” “Inflections,” such as “man’s,” are not names either, since they produce nothing true or false when combined with “is,” “was,” or

“will be”—nor is “walked” a verb; it is an “inflection,” because it signifies additionally “a time outside the present.” In “complex names,” such as “lifeboat,” the parts are significant, but not independently, since, for example, “life” in this occurrence cannot be given an ordinary interpretation (*De Interpretatione*, Chs. 2 and 3). Finally, there are “connections” (*σύνδεσμοί*), words and phrases that “make many things one” (*Rhetoric* III, 12; 4), which seem to include particles, conjunctions, prepositions, and idiomatic phrases of several sorts and which in one passage of doubtful authorship are said to be nonsignificant (*ἄσημοι*) (*Poetics*, Ch. 20). (The “connections” are almost certainly the direct ancestors of Priscian’s “syncategoremata,” which figured prominently in medieval semantics.) This loosely organized classification, vaguely consistent at best, is based on a tangle of semantic and syntactic considerations, but it does contain important advances—for instance, in the treatment of names without bearers and complex names.

Aristotle’s semantics of sentences is concentrated in but by no means confined to *De Interpretatione*, Chapters 4–8. Names have no significant parts and complex names no independently significant parts, but a sentence (*λόγος*) must have independently significant parts. (This is surprising in view of the fact that in a highly inflected language such as Greek there are frequent occurrences of one-word sentences—“I-walk,” “he-walks,” and so on.) “Every sentence is significant—not as a tool but ... by convention,” he maintained (16b33), apparently dissociating his view from Plato’s in *Cratylus* 386D ff. Plato, however, was talking about names, not sentences, and Aristotle here seems to have gratuitously set aside an insight into the semantics of sentences that was later to be developed by the Stoics. Some sentences, such as “prayers” and future contingents, are neither true nor false according to Aristotle, and he set the pattern for nearly all logicians thereafter when he put such sentences aside and attended solely to the always true or false “statement” (*λόγος ἀποφαντικός*).

Aristotle maintained that among the independently significant parts of a statement there must be either a name or an indefinite name and a verb or an inflection of a verb arranged in such a way that the whole “signifies something about something.” It is only in such a combination that there is truth or falsity, and, as Aristotle put it in the early chapters of *De Interpretatione*, it looks as if he took the combination in question to be one of words. In *Metaphysics* 1027b23, however, he said that “falsity and truth are not in things ... but in *thought*; while with regard to *simple* concepts and essences falsity and truth



do not exist even in thought ... but ... the combination and the separation are in thought and not in things,” and he suggested something similar in *De Interpretatione*, Chapter 14 (and elsewhere), as well (cf. Plato, *Republic* 382B). There is no evidence that Aristotle distinguished consistently or clearly between sentences and what later philosophers called propositions or judgments, but such passages indicate at least his sense of the difficulty in locating truth in strings of words, or in a direct relation between strings of words and arrangements of things.

Aristotle seems sometimes to have considered the communicative capacity or public character of a locution as a criterion of its having independent significance. Thus in *Metaphysics* 1006a21 he remarked that if a man “really is to say anything,” he must “say something that is significant both for himself and another”; and in support of his claim that when a verb is uttered by itself it is really a name and signifies something, he noted that on such an occasion “the speaker arrests his thought and the hearer pauses.”

## THE STOICS

The nature of the Stoics’ philosophy of language is the most tantalizing problem in the history of semantics. We know enough of it to say that it was by far the most intricate and probably the most insightful theory of its kind in antiquity and for centuries afterward; but we cannot be certain what its details were, and even its leading principles are sometimes obscured by vague or conflicting testimony. Those Stoics who had most to say about language were, naturally, the logicians, and the difficulty of determining the exact character of what they had to say stems from the fact that none of the many works of the Stoic logicians is extant. The best surviving sources (which date from almost five hundred years after the period of greatest development in Stoic logic and semantics) are Sextus Empiricus, *Outlines of Pyrrhonism*, Book II, and *Adversus Mathematicos*, Book VIII; and Diogenes Laërtius, Book VII. Under these circumstances it is seldom possible to assign a particular doctrine to a particular Stoic, but much of the best of their logic and semantics is very likely to be the work of Chrysippus (c. 280–206 BCE).

Under the Stoic division of philosophy into physics, ethics, and logic, logic was divided into rhetoric and dialectic, and dialectic further divided into an account of language (*περὶ τῆς φωνῆς*) and an account of things signified (*περὶ τῶν σημαινόμενων*). Both these subdivisions contain material relevant to semantics. In their account of language the Stoics distinguished vocal sound generally, “which may include mere noise,” from the sort that is

articulate (*ἔναρθρος*), that is, capable of being embodied in written symbols (*ἐγγράμματος*). Articulate sound, in turn, may be nonsignificant—for instance, “blityri”—or significant (*σημαντικῆ*); but for any articulate sound to be considered a sentence (*λόγος*) it must be significant and a product of someone’s reason (Diogenes Laërtius 7.55–57).

Within that same branch of their dialectic the Stoics recognized five kinds of words and distinguished their semantic or syntactic functions. They were the first who clearly separated (1) names, such as “Socrates,” from (2) appellatives (*προσηγορία*), such as “man.” (Cf. Aristotle’s similar but significantly different distinction in *De Interpretatione*, Ch. 7.) A name “points out a kind proper to an individual,” while an appellative “signifies a common kind.” (3) A verb “signifies a predicate”; (4) a conjunction “binds together the parts of a sentence”; (5) an article (possibly also what would now be called a relative pronoun) serves to “distinguish the gender and number of nouns” (Diogenes Laërtius 7.58). Thus the function of conjunctions and articles is purely syntactic, the semantic function of (proper) names is different from that of appellatives (or common names), and the appellative and the verb—the standard ingredients of the simplest kind of logicians’ sentence—have one and the same kind of semantic function. The appellative occurring in a sentence signifies a subject and the verb a predicate or “something attachable (*συντακτόν*) to the one or more subjects.”

Obviously the division between the accounts of language and of things signified was not exclusive, but the transition from the one account to the other as the Stoics conceived of them may be seen in the claim that all we utter (*προφέρειν*) is sounds, while what we express (*λέγειν*) is matters of discourse (*πράγματα*), or *lekta*—“expressibles” (Diogenes Laërtius 7.57). It is the doctrine of the *lekton* around which the Stoics organized their account of things signified. In its novelty, importance, and difficulty that doctrine overshadows all the considerable remainder of their philosophy of language.

**THE LEKTON.** Probably the clearest introduction of the notion of the *lekton* is the one to be found in these passages from Sextus:

The Stoics ... said that three things are linked together: (1) what is conveyed by the linguistic sign [*τὸ σημαινόμενον*], (2) the linguistic sign itself [*τὸ σημαίνον*], and (3) the object or event [*τὸ τυγχάνον*]. Of these the linguistic sign is the sound—e.g., “Dion”; what is conveyed by the

sign is the matter of discourse indicated thereby, which we apprehend over against and corresponding to our thought (while the barbarians do not understand, although they do hear the sound); and the object or event is the extramental entity—e.g., Dion himself. Two of these are corporeal—viz. the sound and the object or event—and one is incorporeal—viz. the matter of discourse conveyed by the linguistic sign, the lekton. (*Adversus Mathematicos* 8.11–12)

They also say that the lekton comes into being as corresponding to a rational presentation [*λογικὴν θαντασάν*], and that a rational presentation is one presenting something that can be set forth in a sentence. (*Adversus Mathematicos* 8.70)

The kind of lekton associated with the name “Dion” was said to “stand in need of completion,” and the only categories cited for such completable lekta were subjects and predicates. In order to be “set forth in a sentence,” the completable lekta must enter into the composition of a lekton “complete in itself.” The kind of complete lekton regularly associated with a standard subject-predicate sentence was called a statement (*ἀξιωμα*), and truth or falsity was ascribed to it, not to the sentence. Statements naturally received most attention from the Stoic logicians, but they recognized many other varieties of complete lekta as well. The fact that they did so strongly suggests that they had developed other categories of completable lekta too, for most of the other complete lekta cannot be analyzed into subject and predicate. Among the other varieties were commands, prohibitions, yes-no questions, questions requiring more than “yes” or “no,” curses, prayers, doubts (“Can it be that life and pain are akin?”), and quasi statements (“How like to Priam’s son the cowherd is!”) (*Adversus Mathematicos* 8.71–73; Diogenes Laërtius 7.65–68).

Since these are categories of incorporeal lekta rather than of sentences, they cannot be identified with strictly grammatical categories. Moreover, although some of the distinct lekta do correspond to grammatically distinct sentences—for instance, the two kinds of questions—many of them do not. The Stoics’ own example of the kind of lekton called a doubt was expressed in what is grammatically a yes-no question; commands and prohibitions get expressed in declarative as well as in imperative sentences, and occasionally both may be expressed in one and the same sentence, for example, “Abstain from strong drink.” Thus Plutarch reports, in his attack on the Stoics, that “they themselves maintain that those who for-

bid *say* one thing, *forbid* another, and *command* a third. For he who says ‘you ought not to steal’ forbids stealing and commands not stealing at the same time as he says you ought not to steal” (*On the Contradictions of the Stoics* 1037d).

As many as three different complete lekta may, then, be associated with a single sentence, and those lekta are obviously not to be identified as thoughts or intentions on the part of the speaker or hearer. Nor does it seem likely, despite Plutarch’s way of presenting the doctrine, that all the complete lekta associated with a given sentence must be expressed whenever the sentence is uttered. Besides being far-fetched, that requirement would ignore the sense of *expressibility* built into the Stoics’ technical term “lekton.” Instead, the Stoic doctrine seems to be that a number of distinct linguistic jobs—such as stating, commanding, prohibiting—can be performed by means of a single sentence, depending on which of the complete lekta associated with that sentence is actually communicated on a given occasion of its use. Thus the three lekta associated with the example given by Plutarch may be presented as (1) the statement that one ought not to steal, (2) the command not to steal, and (3) the prohibition of stealing. It seems to be a discovery of the Stoics (and their greatest contribution to semantics) that the explication of meaning involves not only the things we talk about and the thoughts we express but also the jobs we do by means of language alone.

## THE EPICUREANS

Of the Stoic semantic triad—linguistic sign, what is conveyed thereby (the lekton), and external object or event—the Epicureans accepted only the first and third, ascribing truth and falsity directly to spoken sentences (Sextus Empiricus, *Adversus Mathematicos* 8.13). This rejection of the lekton is typical of the Epicureans’ mistrust of any doctrine that went beyond the evidence of the senses. Plutarch describes them as “completely doing away with the category of lekta, leaving only words and objects and claiming that the intermediate things conveyed by the signs simply do not exist” (*Adversus Coloten* 1119F), but there is also a vague suggestion that they may have found it convenient to provide “lekta” as dummy referents in one important kind of case. “They deprive many important things of the title of ‘existent,’ such as space, time, and location—indeed, the whole category of lekta (in which all truth resides); for these, they say, are not *existents* [*ὄντα*], although they are *something* [*τινά*]” (*Adversus Coloten* 1116B).

In stating their atomist metaphysics, the Epicureans were of course obliged to use such words as “space” and “time,” and it looks as if they may have clumsily attempted to provide referents for them by associating only such words, or sentences containing such words, with lekta. However, even if they did maintain that there are two kinds of referents for words, real things and lekta, the latter to be invoked only in case the former are unavailable and the words are indispensable, there is nothing of the *Stoic* lekton in their doctrine.

Aside from this putative special use of special lekta, the Epicureans’ philosophy of language seems to have remained remarkably faithful to their fundamental sensationalism. Epicurus (341–270 BCE) had originally stressed the importance of beginning the study of physics (one of the main branches of Epicurean, as of Stoic, philosophy) by ascertaining the ultimate referents (*ὑποτεταγμένα*) of words, “so that our proofs may not run on untested indefinitely nor the terms we use be empty. The primary intent (*ἐννόημα*) of every term employed must be clearly seen and ought to need no explication” (Diogenes Laërtius 10.37–38); and he went on to claim that these ultimate referents must then be “our sensations,” “present impressions,” “actual feelings.” These are always veridical since their immediate causes are the *eidola*, and thus “the agent productive of each of them is always entirely presented and, as being presented, it is incapable of being productive of the presentation without being in very truth as it appears.... Thus the visible object not only appears but actually is as it appears.... The presentations that occur are, then, all true” (Sextus Empiricus, *Adversus Mathematicos* 7.203–204). The square tower in the distance appears round, but its round appearance is itself a physical object, an *eidolon* detached from the tower and impinging on the apparatus of sight. If I say, then, “the tower is round,” I may (and in this case I shall) be mistaken, since the tower is not the immediately presented object. But if I say “the appearance (or presentation) of the tower is round,” I cannot be mistaken (at least not in that same way). Although this is a move in the direction of protocol sentences, it does not rest on a distinction between sense-datum and physical-object sentences, since for the Epicureans the protocol sentence was only a more correctly framed physical-object sentence.

No full account of this Epicurean reductivism is extant, but its principle is clearly operative not only in their physics but in their ethics as well, where one pervasive maxim for the avoidance of fear is to reduce the mysterious (for example, in natural phenomena) to what is

actually presented and to describe it in terms precisely associated with the features of the actual presentation (see, for instance, Diogenes Laërtius 10.78 ff.).

Epicurus’s followers evidently took the nature of the relation between words and sensations as a major topic for psychological theory. The notion of *prolepsis* is at the center of the Epicurean psychology, and in at least one of its many guises *prolepsis* seems to be the act of associating a word with a *typos*, or outline left in the mind as the result of repeated similar presentations. One example of *prolepsis* is the identification “such and such a thing is a man”—“for no sooner is the word ‘man’ uttered than we think of the *typos* of man in accordance with the *prolepsis*, the senses having led the way. As a result, the immediate referent of every name is apparent.... Nor would we have given a name to anything if we had not first come to know the *typos* of it in accordance with *prolepsis*” (Diogenes Laërtius 10.33).

The *typos*, then, is the immediate referent (*τὸ πρῶτως ὑποτεταγμένον*) of every name. When a name is used and understood, an act of *prolepsis* at once brings the corresponding *typos* to mind. (Since nothing but sensation can produce a *typos*, the need for some other sort of referent in the case of words such as “space” and “time” is apparent.) If this was indeed the core of the Epicurean semantics of words, it must be judged inferior to many other theories of its kind in antiquity.

Epicurus himself and the Epicureans generally had a good deal to say about the origin of language, and what they said usually makes better sense than most such accounts in antiquity. Lucretius (99–55 BCE) is especially good on this topic, which he treated at some length in Book V of his poem. Among his more novel and interesting achievements is an extended series of arguments against the theoretical possibility that language (as distinguished from *a* language) might have been *invented* (*De Rerum Natura* 5.1041 ff.).

## THE MIDDLE AGES

### ST. AUGUSTINE

Most of what St. Augustine (354–430) had to say about language and meaning was said not for its own sake but in support or elucidation of some theological doctrine. Partly for that reason, perhaps, his semantic doctrines had less effect on philosophy of language in the Middle Ages than might be expected, considering his enormous influence on medieval philosophy in general.

The short treatise *Principia Dialecticae* (probably written around 384, when Augustine was a professor of rhetoric) contains what may be the only instance of a semantic inquiry pursued by Augustine without a motive. In it he distinguishes four principal semantic elements: (1) the word (*verbum*), a spoken articulate sound, classifiable as a vocable of some language; (2) the expressible (*dicibile*), “whatever is sensed in the word by the mind rather than by the ear and is retained by the mind”; (3) the ordinary use of the word (*dictio*), (opposed, for instance, to the use of the word as a sign for itself), which involves “both [1] the word itself and [2] that which occurs in a mind as the result of the word” when the word occurs “not on its own account but on account of something else that is to be signified”; (4) the signified thing (*res*), which may be “something understood, or sensed, or inapprehensible”—the last category reserved for, for instance, God and formless matter (Ch. 5). Of these four elements, (2) and (3) together seem to represent different aspects of the Stoic lekton; but whatever their origin, their inclusion here indicates a level of sophistication in semantics that was not to be attained again for at least eight hundred years.

Chapter VII of the *Principia Dialecticae* is devoted expressly to the “import” (*vis*) of words. In it Augustine maintains that “the import of a word is that whereby the extent of its efficacy is recognized [*qua cognoscitur quantum valeat*], and it is efficacious to the extent to which it can affect a hearer.” Import is a broader notion than signification and includes several sorts of effects a given word may have because of its sound, its degree of familiarity to the hearer, its degree of admissibility into polite conversation, its being recognized by the learned hearer as a dactylic foot or as some particular part of speech, and so on. The paradigm case of signification is described as occurring “on an occasion when a sign has been comprehended through a word [and] the mind regards [*intuetur*] nothing other than that very thing the sign it comprehends is a sign of. Suppose, for example, that Augustine has been named and someone to whom I am known thinks of nothing other than myself, or that some other man named Augustine comes to mind if the name happens to be heard by someone who does not know me but knows that other man.” The most remarkable and apparently novel features of this brief account are (a) the extension of the notion of meaning in Augustine’s doctrine of “import” and (b) the orientation of his account of meaning around the effects words have on their hearers. The remainder of the treatise deals with simple and conjoined words and sentences, etymology, and various types of ambiguity and obscurity.

The longest of Augustine’s discussions of semantic questions occurs in the dialogue *De Magistro* (389), which is designed ultimately to support the Augustinian doctrine of “divine illumination” as the sole genuine source of truth. Thus the first 11 chapters are supposed to show that “we learn nothing through those signs that are called words” (Ch. X), while chapters XI–XIV develop the thesis that Christ, the truth, teaches us inwardly while men by their use of outward signs merely prompt us to raise questions. The argument in support of the negative conclusion is an outstanding example of overemphasis on the word as the unit of signification. “When words are uttered, either we know or we do not know what they signify. If we know, then we do not learn but are reminded. If, on the other hand, we do not know, then we are not even reminded (though we may be prompted to ask)” (Ch. XI). Therefore, “we learn nothing through those signs that are called words”—as if one’s knowing what the words mean in “armadillos are mammals” precluded one’s learning anything through hearing that sentence uttered. At this crucial juncture in the dialogue Augustine’s ulterior motive seems to have distorted his judgment.

Perhaps the most interesting point in the early chapters of *De Magistro* is one that bears on the best-known Augustinian passage on language, the description of his learning to speak in *Confessions* (397), 1.8, made famous by Ludwig Wittgenstein’s use of it in *Philosophical Investigations* (Sec. 1). The passage in the *Confessions* can hardly be considered a theoretical statement at all, since Augustine’s main aim in it is to describe a milestone on his descent “into the stormy fellowship of human life,” but it does contain a brief, uncritical account of one way in which a child might be shown “the things of which words are signs.”

That this account cannot be considered important in the context of Augustine’s own views on language is plain from the fact that he had already criticized just such an account on theoretical grounds in *De Magistro*. In an attempt to refine the original suggestion of the dialogue that a sign cannot be a sign “unless it signifies something” (Ch. II), Augustine asks to be shown “that one thing itself, whatever it is, which is signified by these two words,” *ex* and *de*, Latin prepositions there taken to be synonymous. After several obviously unsuitable suggestions, the tentative conclusion is reached that not only in these problematic cases but also in every case of attempting to show “the thing signified,” all that can be shown is further signs, such as other words, pointing, pantomiming. This criticism is of course not the same as Wittgenstein’s, nor is it

particularly far-reaching in its own right, since it is soon modified to allow that we can in certain cases display the very thing signified without the use of further signs—for instance, if the thing signified is something we are able to do (such as walk) and we are not in the act of doing it when asked to display the thing signified (Ch. III).

A rather fully developed semantic theory appears in the *De Trinitate* (399–419), especially in 9.7–12 and 15.10–16, although it is presented no more for its own sake than is the theory in *De Magistro*. The theory appears as the explanatory half of an ingenious analogy designed to clarify (a) the relation between the First and Second Persons of the Trinity, (b) the two natures of the Second Person, and (c) the identification of the Second Person as the Word. The analogical points may be ignored for present purposes and the semantic doctrine sketched as follows. “Word” has at least two senses. “In one sense, those things are called words that occupy intervals of time with syllables, whether they are pronounced or only thought. In another sense, everything that is known is said to be a word impressed on the mind as long as it can be brought out of memory and defined” (9.10). (Augustine actually introduces a third sense involving the love of what is known, but it seems pointless except for purposes of the analogy.) The second kind of “word,” which Augustine describes more generally as a “locution” when the demands of the analogy are not uppermost in his mind, occupies the central position in the doctrine. “The word that sounds outwardly is a sign of the word that gives light inwardly, and the name ‘word’ is better suited to the latter; for what is uttered by the mouth of the flesh is the articulate sound of the word [*vox verbi*]; ... [thus] our word becomes an articulate sound ... by taking on [articulate sound], not by consuming itself so as to be changed into it” (15.11).

The doctrine of the inward locution sometimes bears a striking resemblance to Plato’s doctrine of ideal names in the *Cratylus*, although a direct historical connection seems unlikely. “For of necessity, when we say what is true—i.e., say what we know—the knowledge itself, which we retain in memory, gives birth to a word that is altogether of the same kind as the knowledge from which it is born. For the thought formed by the thing that we know is a word that is neither Greek nor Latin nor of any other language. But since it is necessary to convey it into the knowledge of those with whom we speak, some sign is adopted by which it is signified” (15.10; cf. *Sermo* 225.3). According to this doctrine, then, it seems that one’s saying “armadillos are mammals” embodies in sounds one’s inward locution to that effect, which itself

differs from one’s knowledge that armadillos are mammals only in being brought out of memory into conscious thought. Augustine sometimes suggests that the inward locution, then, is not itself verbal; words used in the mind are not essentially different from words outwardly pronounced, as Augustine’s first division claims. Indeed, the inward locution is evidently less a mental entity than the state of consciousness into which a mental entity, namely, a known truth, must be brought if it is to be given verbal expression.

## BOETHIUS

As an original contributor to semantics, Boethius (480–524) is much less interesting than Augustine. Since, however, his translations and commentaries constituted the sole source of Aristotelian logic for the medievals until the twelfth century, Boethius’s influence over the development of semantics in the Middle Ages is powerful where Augustine’s is slight.

Most medieval semantic theories take as their starting point Boethius’s translation of the rudimentary account in Aristotle’s *De Interpretatione*, Chapter 1. No doubt the traditional misreading of those passages during and after the Middle Ages is largely the result of the fact that in his otherwise faithful rendering Boethius obliterated the Aristotelian distinction between symbols and symptoms, translating both *σύμβολα* and *σημῖα* as *notae*. Another of the principal difficulties in Aristotle’s account—the apparent interposition of “mental modifications” between words and things—had been discussed at least as early as the third century by Alexander of Aphrodisias, whose confusing resolution of the difficulty was transmitted to the medievals in Boethius’s second commentary on the *De Interpretatione*. Alexander had asked whether Aristotle’s account forces us to consider the mental modifications as names of things. In order to avoid that consequence he had developed the view that although “a name is imposed on a thing” and “although spoken words are names of things, nevertheless we use spoken words not in order to signify things, but in order to signify those mental modifications that are produced in us as a result of the things. Therefore, since spoken words are uttered for the purpose of signifying those entities, he [Aristotle] was right to say that they are primarily the signs [*notas*] of those entities” (413A–B; all references in this section are to *Patrologia Latina*, edited by J. P. Migne, Vol. 64).

Perhaps the most influential doctrine (at least in the late Middle Ages) that can be traced directly to Boethius’s treatment of *De Interpretatione*, Chapter 1, is that of the

three discourses: written, spoken, and mental. Citing Porphyry (c. 233–c. 305) as his authority, Boethius reported that “among the Peripatetics there were said to be three discourses [*orationes*]—one written in letters, another uttered in speech, and a third put together in the mind. Now if there are three discourses, the parts of discourse are no doubt likewise threefold; for since the noun and the verb are the principal parts of discourse, there will be some nouns and verbs that are written, others that are spoken, and still others that are silent and employed by the mind” (407B–C). Here, as in his transmission of the Aristotelian account itself, the vagueness of Boethius’s presentation is as important historically as its content. Are there two completely different sets of nouns and verbs, one for writing and one for speech? And is this mental discourse nothing more than silently running over a sentence in Latin or English, Or is it a nonverbal operation, reminiscent of Augustine’s “inward locution”? The fact that mental discourse is said to have nouns and verbs of its own suggests the former view, if either; but since Aristotle had maintained that the mental modifications were the same for all (regardless of their native tongue), and since Boethius offers this doctrine of the three discourses in explanation of Aristotle’s account, there is some basis for the second view as well. These were among the difficulties discussed in the medieval development of the doctrine.

The medieval distinction between words of first and second “imposition,” a genuine prefiguring of the twentieth-century distinction between object language and metalanguage, also has its roots in Boethius’s transmission of older doctrines. In his commentary on Aristotle’s *Categories* he presents the distinction very much as he found it in Porphyry’s *Expositio* of the same work (A. Busse, ed., pp. 57–58). “The first imposition [*positio*] of a name is made with respect to the *signification* of the word, the second with respect to its *form*” (159C). Thus, whenever some extralinguistic entity is called a man, it is a case of first imposition. “But when the word ‘man’ itself is called a noun, no reference is made to the signification of the word [Boethius has ‘noun’], but to its form, in virtue of which it admits of inflection by means of [grammatical] cases” (159B–C). Thus “noun” in its ordinary use is a word of second imposition.

In this primitive form the distinction seems to apply only to the grammarian’s kind of interest in discourse. Boethius, however, took the position that “the whole art of logic is concerned with discourse” (161C–D). How does the philosopher’s interest in language differ from the grammarian’s? Very much as the economist’s interest in

money differs from the numismatist’s, for Boethius compares the signification of a word to the buying power of a coin and its grammatical form to the “bronze stamped with a design.” Consider “an utterance that designates nothing, such as ‘gargulus.’ Although the grammarians, considering its form, contend that it is a noun, philosophy does not recognize it as a noun unless it is imposed in such a way as to designate a conception belonging to a mind (in which same way it can signify some real thing)” (408C–D). Apparently, then, second imposition needs to be more broadly conceived, or a philosopher’s kind of second imposition must be added to the kind described by Boethius. The resolution of such difficulties was among the goals of the later doctrine of the impositions and “intentions” of words.

By far the most influential of Boethius’s bequests to the Middle Ages was his formulation of the problem of universals in his second commentary on Porphyry’s *Isagoge*. Needless to say, a great many semantic issues were discussed in the long controversy over universals, and a few of the more important ones will be noted below. Boethius’s formulation of the problem, however, was oriented around questions of metaphysics rather than of semantics and so may be passed over here.

## ST. ANSELM

One of the semantic problems recognized by the early medievals in the few logical works of Aristotle available to them was the problem of paronyms, or denominatives. Its principal source is the following passage in the *Categories*, Chapter 8 (10a27 ff.). “These, then, that we have mentioned are qualities, while things called paronymously because of these or called in some other way from them are qualified. Now in most cases, indeed in practically all, things are called paronymously, as the pale man from paleness, the grammatical from grammar, the just from justice, and so on.”

St. Anselm (1033–1109) remarks at the end of his dialogue on denominatives—*De Grammatico*—that the semantics of denominatives was a favorite topic among dialecticians of the eleventh century, evidently because of the difficulty of developing a satisfactory account of denominative words that occur both as concrete nouns and as adjectives. (Anselm’s chief example is *grammaticus*, but because the English word “grammatical” is not a denominative of this sort, “illiterate” will be used here.) Thus, the opening question of Anselm’s dialogue is whether “illiterate” signifies a substance or a quality. This seems to be a narrow, perhaps artificial problem, but

under his characteristically ingenious treatment it leads to results of general importance.

The superficially most plausible solution to the problem is that such a word sometimes signifies a substance—as in “not every illiterate is stupid”—and sometimes a quality (illiteracy)—as in “not every illiterate person is stupid.” This solution is shown to fail, however, at least in its second half, for if we tried to use “illiterate” alone in speaking about the quality—as in “illiterate is a deplorable condition”—“not only the grammarians would be upset, but even the peasants would laugh” (Ch. XI). “Illiterate,” we must recognize, “does not signify a person and illiteracy as a unit [*ut unum*] but signifies illiteracy directly [*per se*] and a person indirectly [*per aliud*]” (Ch. XII). Another way to put the distinction between the two kinds of signification is to say that “illiterate” is *significative* of illiteracy and *appellative* of a person. “I now describe a name as appellative of each thing itself that is called [*appellatur*] by that name in the speaker’s usage; for there is no speaker’s usage in which ‘illiteracy is illiterate’ occurs, ... but rather ‘the person is illiterate’” (Ch. XII).

The remainder of the dialogue refines and generalizes this account in the course of dealing with various objections to it. Anselm’s most original and important contributions seem to be those developed mainly in the last two chapters (where the discussion centers around *albus*—“white”—rather than *grammaticus*). The Master of the dialogue has suggested that “white” signifies (rather than appellates) nothing but being in possession of whiteness (*habens albedinem*). This is disturbing to the Student, who feels the need of a signified *thing*. “White,” he is willing to grant, “does not determinately signify this or that possessing entity, such as a body,” but he wants to insist that it “indeterminately signifies something possessing whiteness.” His principal argument is that “‘white’ signifies either something possessing whiteness or nothing; but one cannot conceive of nothing as possessing whiteness; therefore it is necessary that ‘white’ signify something possessing whiteness” (Ch. XX).

In reply Anselm takes the position that while it may always be the case that what is signified somehow depends for its being on some real thing, it cannot always be the case that what is signified is a thing. His arguments for this position display an interesting use of the principle of substitutivity.

If “white” signified a thing at all, it would signify something white. Now the signification of a word is what its definition presents, and what is presented by the definition may be substituted for the word itself. “So wherever ‘white’ is used it is taken correctly as ‘something

white” (Ch. XXI). Then “Socrates is white” may be rewritten as “Socrates is something white.” But “wherever ‘something white’ is used it is also correctly said twice—‘something something white’—and wherever it is said twice, there also three times, and so on indefinitely” (Ch. XXI). Thus the plausible “Socrates is something white” would become the nonsensical “Socrates is something something white” and would ultimately lose all semblance of a statement.

Instead, in “Socrates is white,” “white” appellates something white—Socrates himself—but what it signifies is being in possession of whiteness. Nor will it do to introduce a signified thing at this point, for if we take something in possession of whiteness to be what “white” signifies, we shall have to grant that something in possession of whiteness is that which is white. “If, therefore, ‘white’ is ‘that which is white,’ it is also ‘that which is that which is white’; and if it is that it is also ‘that which is that which is that which is white,’ and so on indefinitely” (Ch. XXI). The nonsense-engendering substitutions cannot be made within “being in possession of whiteness,” however, since the denominative “white” does not itself occur in it. Thus “it is clear enough that ‘white’ does not signify something in possession of whiteness ..., but only being in possession of whiteness—i.e., [the categories] *quality* and *possession* [and not the category *substance*]—and quality and possession by themselves make up no *something*” (ibid.). This argument is described as holding good for all single words that, like “white,” signify “a plurality [of categories] out of which no one thing is made up” (Ch. XXI).

Although the special consideration of denominatives apparently lost its vogue soon after Anselm, many of the problems dealt with in his *De Grammatico* remained current and can be found two centuries afterward at the center of the theory of the properties of terms (see below).

## ABELARD

The extensive logical writings of Peter Abelard (1079–1142) are best known for the theory of universals developed in them. That theory is important in the history of semantics because (a) it explicitly approaches the problem of universals as a semantic rather than a metaphysical problem and because (b) in doing so it introduces many of the elements of the semantic theories developed by the terminist logicians of the thirteenth and fourteenth centuries.

Regardless of how the problem of universals is approached, it involves a consideration of the semantics of words, especially of common names. Nevertheless,

many of the countless medieval theories, in their preoccupation with the Porphyrian-Boethian questions about the existential status of genera and species, slighted or ignored the semantic issues. Abelard, on the other hand, began by adding a new semantic question to the three traditional metaphysical questions. “Could a universal consist of the signification of a concept [*significatio intellectus*] when the things named were destroyed, as [in winter] when there are no roses to which the name ‘rose’ is common?” (*Logica “Ingredientibus,”* edited by B. Geyer, p. 8).

Having associated universals with words, Abelard asked “whether they are associated *only* with words or with things as well” (p. 9). Applying the Aristotelian criterion *predicability of more than one thing*, he showed in a series of elaborate arguments that a universal cannot be identified as (1) a single thing or (2) a collection of things (pp. 10–16). His principal objection really avoids the issue of whether or not it makes sense to speak of a thing or a collection as predicable at all and concentrates instead on the impossibility of predicating a thing or a collection of more than one thing. Thus “it remains for us to ascribe universality of that sort to words [*vocibus*] alone” (p. 16). As Abelard came to realize, words considered as utterances or inscriptions are themselves things. Accordingly he eventually distinguished between utterances [*voces*] and words [*sermones*] and organized his theory of universals around words in this strict sense—*sermo = vox + significatio*—which he described as products of human arrangements rather than mere natural effects (*Logica “Nostorum Petitioni Sociorum,”* edited by B. Geyer, p. 522).

The only kind of word to which universality can conceivably be ascribed is the kind of word apparently predicable of more than one thing, that is, a common name in the nominative case. But that ascription cannot mean that the common name has some universal thing as its bearer, for, as he had shown, “universal thing” is a contradiction in terms. Nor can some particular thing be picked out as its bearer, for although it *may be* Socrates alone of whom the statement “a man is sitting in this house” is true, we cannot *infer* from it that Socrates is sitting in this house (p. 18). These considerations led Abelard to base the ascription of universality not on what the words name (*nominare*)—for example, “man” names each and every individual thing that is a man—but on their “mode of signification”; for although they name things that are discrete, they do so not “discretely and determinately” but “confusedly” (p. 29).

Abelard’s explanation of this notion of confused naming, which was to play an important part in thirteenth- and fourteenth-century theories of the properties of terms, seems incomplete but runs along the following lines. “To signify is to establish [*constituere*] a concept” (p. 136), and “when I hear ‘man’ ... I do not recall all the natures or properties that are in the things subject [to that name]; instead, as a result of [hearing] ‘man’ I have a conception of animal, rational, and mortal, though not of subsequent accidents as well, [a conception that is] confused rather than discrete” (p. 27). Thus Abelard’s answer to his additional semantic question is a qualified “yes.” In winter the name “rose” lacks universality in that there are no things of which it is predicable, that is, “it is devoid of nomination” (*nominatione*). “Nevertheless, it is still significative then in virtue of the concept [*ex intellectu*]; ... otherwise there would not be the proposition ‘no rose exists’” (p. 30).

Other medieval theories of universals, such as William of Ockham’s, center on semantic doctrines; but Abelard’s “sermonism” was perhaps the most important medieval influence on the development of semantics during the succeeding two centuries of the high Middle Ages. Topics and terminology remained relatively stable in that remarkable period in the history of semantics, although many philosophers and every logician contributed to the discussions. For that reason the remaining material on the Middle Ages is oriented mainly around *topics* in medieval semantics, and no attempt is made to mention every man who discussed them.

## IMPOSITIONS AND INTENTIONS

The pervasive medieval distinctions between two levels of signification have attracted some attention in the twentieth century because of their resemblance to the object language–metalanguage distinction. Historically there were two such distinctions, both based on the observation, found already in Porphyry, that while some signs signify nonsigns, others are signs of signs.

The original distinction was drawn with respect to conventional signs, specifically, with respect to names (nouns and adjectives) in a natural language. Such signs acquired their signification only as a result of having been imposed by the users of the language. The primary, or first, imposition was on extralinguistic entities, and names such as “man” and “white” were classified as names of first imposition. As the language developed, other conventional signs were imposed on conventional signs as such; thus, names such as “noun” and “plural” are of second imposition. Names such as “utterance” and “mark”



do signify conventional signs, but not as such (since there are of course nonsignificant utterances and marks); they are therefore names of first imposition. Most medieval logicians, presumably in avoidance of an infinite regress, were careful not to define names of second imposition as names of names of first imposition. Thus “name of second imposition” is itself a name of second imposition.

But even those who, like Abelard (*Logica “Ingredientibus,”* edited by B. Geyer, p. 112), did define second imposition in terms of first seem never to have recognized a “third” imposition. (The imposition distinction, therefore, cannot reasonably be described as prefiguring a hierarchy of types.) The use made of the imposition distinction was apparently rather meager. Aristotle’s categories were, for example, often said to be names of first imposition, while the subject matter of his *De Interpretatione* was described as names of second imposition. The distinction, although it was refined and discussed well into the fifteenth century, seems to have acquired what importance it had mainly from its connection with the later and better known of the two distinctions between levels of signification.

Concepts in their capacity as natural signs were called intentions and described in the doctrine of the three discourses as mental terms. It was only natural, then, to distinguish levels of signification among intentions as among conventionally significant extramental terms. This distinction, probably stemming from Avicenna (see Carl Prantl, *Geschichte der Logik im Abendlande*, 2.328), classified as first intentions all those naturally significant of entities other than intentions as such, while those that did naturally signify intentions as such were second intentions. The concept *humanity* is of course a first intention, but so is the concept *mental entity*. The concepts of the predicables—*genus*, *species*, *differentia*, *property*, *accident*—are second intentions, as is the concept *predicable* itself; no “third” intentions were ever recognized.

Thus first and second intentions and impositions were fundamentally parallel distinctions in separate domains. However, their development in the thirteenth and early fourteenth centuries was complicated (and sometimes confused) by two factors. First, there were, of course, extramental terms imposed on first and second intentions, such as *humanity* and *genus*. Such names were all of first imposition (since no intention was a conventional sign), but they were sometimes further described as names of first or second intention. Second, even more complicating was the fact that the first and second intentions themselves were considered to be terms in mental

propositions. Thus, while in the written proposition “*animal* is a genus” the subject and predicate terms are both of first imposition, in the corresponding mental proposition that *animal* is a genus the subject term is a first intention and the predicate term is a second intention.

Of the two distinctions between levels of signification, the intention distinction had much more philosophical importance. The confusing interrelations of the two distinctions are perhaps best exhibited in William of Ockham (d. 1349), particularly in *Summa Logicae*, 1, 11–12. Logicians after William—for instance, Albert of Saxony (d. 1390), Pierre d’Ailly (1350–1421), Paul of Venice (d. 1428), and Paul of Pergula (d. 1451)—exhibited a tendency to simplify them by reverting to the treatment of impositions and intentions as strictly separate, parallel distinctions. Postmedieval scholastics—for example, John of St. Thomas (1589–1644)—were inclined to apply the intention distinction indifferently to extramental as well as to mental terms and to ignore the imposition distinction; it is in this simplified form that the “medieval” distinction between levels of signification is usually discussed in recent literature.

### SCIENTIA SERMOCINALIS

Almost everything genuinely novel in medieval logic is to be found in the theories of the properties of terms and of the functions of syncategorematic words developed by the logicians of the high Middle Ages. One reason why logic set off along that line of logicosemantic inquiries is that medieval logicians, especially through the formative period ending about 1250, thought of their subject as the science of language (*scientia sermocinalis*).

That classification itself marked a break with the Aristotelian-Boethian tradition in that it was precise where the tradition had been vague. The notion of predication was unquestionably an essential part of the subject matter of logic, but Aristotle and Boethius had treated it in ways that often suggested that predicates might be extralinguistic and even extramental entities. This crucial vagueness, which was to some extent also the source of the medievals’ concern with universals, left open the possibility that logic might be essentially a science of reality, resembling or subsumed under metaphysics.

However, in the earliest complete European logic we have after Boethius—the *Dialectica* of Garland the Computist (d. before 1102)—that possibility was already noted and explicitly ruled out. Predication, Garland maintained, occurs only in a proposition, and the only constituents of propositions are utterances; thus, only

utterances may be predicated. The five predicables (genus, species, differentia, property, accident), the elementary subject matter of medieval logic, are, in virtue of being predicables, utterances and no more; and the ten categories (substance, quality, and so on) are likewise categories of utterances only—for instance, noun and adjective (*Dialectica*, edited by L. M. de Rijk, p. 3).

The attempt to establish logic as a science of linguistic entities only may be called sermocinalism. During the years 1150 to 1250, when medieval logic was acquiring its distinctive character, sermocinalism held undisputed sway as the philosophy of logic, but it did so in the refined and strengthened form given it in the writings of Abelard. Garland had attempted to make utterances (*voces*) the elements of logic, which he thought of as the science of language. Abelard, recognizing that utterances are physical events that are, as such, of no interest to logicians, replaced the overly simple utterance with what he called the *sermo*, defined as the utterance taken together with its signification. Logicians in the second half of the twelfth century seem to have been unanimous in their adoption of this refinement. An anonymous *Dialectica seu Logica* supported the rejection of utterances as the elements of logic with the following interesting argument, somewhat reminiscent of Aristotle's doctrine of complex names. "Some utterances are significant; some are not.... This division ... is exhaustive but seems not to be exclusive, since the same utterance may be both significant and not significant. For example, the utterance 'king' [*rex*] is significant as a word, but since it is also part of a word, a syllable of a word—as in 'smoking' [*sorex*, shrew]—it is in that case and on that account not a significant utterance" (Martin Grabmann, *Bearbeitungen*, Berlin, 1937, p. 30).

Having more precisely identified the elements of logic as linguistic entities, Abelard suggested that logic as the science of language should determine significations on the basis of the application of utterances, determining the proper application of utterances on the basis of the investigations of the natural sciences (*Dialectica*, pp. 286–287). One reason for this suggestion seems to have been his concern with propositions true *gratia terminorum*, analytically true on semantic rather than on syntactic grounds—for instance, "if there is paternity, there is filiation" or "if it is a body, it is corporeal" (see pp. 284–286).

To most medieval philosophers Abelard's emphasis on the importance of signification as well as of utterance might have suggested that mental entities of some sort were to be considered the elements of logic. He explicitly rejected this possibility, however, and in doing so made

his most important contribution to sermocinalism. He argued that a proposition true *gratia terminorum* could not be verified by an appeal to the status of mental entities. "When we say 'if it is man it is animal,' if we refer to the connection of the *understanding* of the propositions, as if we were concerned with the *concepts*, there is no truth to the conditional, since the one *concept* may occur entirely without the other" (p. 154). What we are concerned with, Abelard maintained, is the connection between the term *animal* and the definition of the term *man*—namely, the inclusion of the term *animal* within the string of terms making up the definition of the term *man*. As a result of this move, sermocinalism was directed not only against the notion of logic as a science of reality (*scientia realis*) but evidently also against the notion of it as the science of reason (*scientia rationalis*).

The philosophy of logic that eventually challenged sermocinalism concentrated its opposition on this last point. Since it was explicitly drawn from the philosophy of Avicenna, the rival doctrine may conveniently be called Avicennianism. Although as many of Avicenna's writings as were available to the medievals had been translated into Latin around the middle of the twelfth century, Avicennianism as a philosophy of logic seems not to have come into prominence until Albert the Great (1193–1280) adopted it around the middle of the thirteenth century. By that time, however, medieval logic was firmly committed to its distinctive line of development as the *scientia sermocinalis*. As a result, the main impact of Avicennianism as an alternative to sermocinalism was felt less on the work of the logicians than on the metaphysicians' discussions of the nature of logic.

The central doctrine of Avicennianism is presented in the frequently quoted passage from Avicenna's *Philosophia Prima*: "The subject matter of logic, as you know, is intentions understood *secondarily*, which are applied to intentions *primarily* understood" (I, 2, f70vA). Logic was the science of reason, Avicenna claimed, for "the relation of this doctrine [logic] to internal thought, which is called internal speech, is just like the relation of grammar to outward signification, which is called speech" (*Logica* f3rA). Thus grammar, not logic, was the sermocinal science, according to Avicennianism, and the rise of speculative grammar that was to follow may in part be attributed to this point of view.

## THE PROPERTIES OF TERMS

Until about the middle of the twelfth century the subject matter of medieval logic was drawn from Aristotle's *Categories* and *De Interpretatione*, together with a set of

books by Porphyry and Boethius that were centered more or less closely on those two books of Aristotle. Later in the Middle Ages this collection of books, or the kind of logic these books contained, became known as *logica vetus*, the old logic. When the remaining four books of Aristotle's Organon began to circulate in western Europe during the twelfth century, they, or their contents, became known as *logica nova*, the new logic. The only completely new kind of material in the *logica nova* was the treatment of fallacy in Aristotle's *Sophistical Refutations*, which excited a tremendous interest in *sophismata*, fallacies resulting from the misuse of or natural ambiguities in various devices of ordinary discourse. Largely because of this lasting interest, medieval logicians of the late twelfth and early thirteenth centuries gradually developed an original logicosemantic inquiry. In order to distinguish this genuinely medieval contribution from Aristotle's contributions to logic, thirteenth-century philosophers began to speak of it as the *logica moderna*, lumping the *logica vetus* and *logica nova* together as *logica antiqua*. Perhaps the *logica moderna* was aimed originally at nothing more than providing ad hoc rules of inference to cover problematic locutions in ordinary discourse, but, although it retained that aim throughout its three-hundred-year history, its principal aim soon became the development of a reasonably general account of the different ways in which words are used to stand for things and to operate on other words.

The earliest known fully developed productions of the "modernist" or "terminist" logicians are the logical treatises of William of Sherwood (d. 1266/1272), Peter of Spain (d. 1277), and Lambert of Auxerre (fl. 1250), evidently written at Paris about the middle of the thirteenth century. At that time the *logica moderna* seems to have been thought of as having two branches, an account of "the properties of terms" (*proprietas terminorum*) and an account of the signification and function of "syncategorematic words" (*syncategoremata*). The two branches naturally differed in detail, but both accounts employed the same principles of explanation and had the same aims. Most nouns, pronouns, verbs, participles, and adjectives were considered to be categorematic words, words capable of serving as terms (that is, as subjects or predicates); and the syncategorematic words were those which can occur in a statement only together with categorematic words. The two branches of the *logica moderna* were thus theoretically exhaustive of the kinds of words occurring in various roles in statements.

The modernists of the thirteenth century regularly recognized four properties of terms: (1) signification—

the word's meaning, broadly conceived, or the range of conventional uses of the word (a property of every categorematic); (2) supposition—the conventional interpretation of a word on a particular occasion of its use, a modification of its signification resulting from its syntactic context, if any, and other considerations (a property only of nouns, pronouns, and "substantive expressions," that is, other categorematics employed as substantives and particularly as subjects); (3) copulation—virtually the same as supposition, except that it is a property only of verbs, participles, and adjectives, especially when they occur as predicates; (4) appellation—"the present correct application of a term" (Sherwood), a property only of nouns, adjectives, and participles; for instance, in 2004 Chicago was an appellatum of "city" but Nineveh was not.

Obviously these four properties are not on an equal footing. The supposition, copulation, or appellation of a term was considered a function of its signification; Vincent of Beauvais (d. 1264) even designated signification the genus of which the other three are species (Carl Prantl, *Geschichte der Logik*, Vol. III, p. 83, n. 319). Moreover, copulation and appellation are of distinctly secondary importance. By the middle of the fourteenth century only signification and supposition were regularly recognized as properties of terms, and throughout the history of the *logica moderna* it was the supposition (*suppositio*) of terms on which the inquiry centered. For that reason the best way of quickly acquiring a broad but accurate idea of the modernists' account of the properties of terms is to examine their divisions of supposition. (The recognition and treatment of the divisions of course differed from one modernist to another, but the following selection includes all the major divisions and many of the more interesting minor divisions.)

The supposition of a term was divided initially into *proper* and *improper* supposition. A term had improper supposition when it was used figuratively, and several varieties of improper supposition were distinguished: antonomastic, synecdochic, metaphoric, ironic, and metonymic. The proper supposition of a term was divided into *formal* and *material* supposition, the latter being the use of a term to refer to itself, either as type or as token—for instance, "man is a noun," "man is an animal" is a true statement," "man is a monosyllable," "man has three letters." Formal supposition was *personal* if the term was used to refer to individuals bearing the form signified by the term, *simple* if the reference was to the form itself, as in "man is a species." The initial division of personal supposition was sometimes based on the divi-

sion of terms as *common* or *discrete*, depending on the possibility of using them to refer to more than one individual at a time. Thus, the subjects of the statements “*Socrates* is running” and “*that man* is running” have discrete personal supposition.

The portions of supposition theory dealing with the divisions of the personal supposition of common terms were more fully developed than the rest, not only because they were intrinsically more interesting but probably also because they provided the most points of contact between the *logica antiqua* and the *logica moderna*. In those portions of supposition theory, far more than in the others, the emphasis lay on the application of the theory to the evaluation of inferences, especially such inferences as involved Aristotle’s four categorical propositions (or near relatives of them) but could not be adequately evaluated within the *logica antiqua*.

A common term was said to have determinate personal supposition when it was used to refer to some one individual without identifying the supposed individual, as in “a *man* is running” or in “some *man* is running.” A statement including a “distributive sign” (such as “every” or “no”) was, on the other hand, bound to include one or more common terms having *confused* personal supposition, terms used to refer to more than one individual at once or to one individual many times (as in “every man sees a *man*,” where the second occurrence of “man” has confused supposition even if it is being used to refer to only one individual). If the confused supposition included each and every individual bearing the form signified by the common term, it was designated *distributive*, as in “every *man* is an animal,” “no *man* is an ass.” If the confused supposition did not plainly include that totality, it was designated *merely confused* (*confusa tantum*), as in “every man is an *animal*,” “every man sees a *man*.”

The modernists observed that in many cases of distributive confused supposition it was possible to make a “descent” under the term having such supposition, instantiating as in “every man is an animal, therefore this man is an animal”; “no man is an ass, therefore no man is this ass.” They described such cases as *mobile* but paid at least as much attention to the *immobile* distributive supposition produced by the use of “exclusives” or “exceptives” together with distributive signs. Thus, in “only every man is running” the distributive supposition of the common term is immobilized by the exclusive “only,” so that one cannot infer “therefore only this man is running.” It was also recognized that the inclusion of the distributive sign within the scope of the exclusive or

exceptive was not always dependent on their relative positions in the statement, since from “every man except Socrates is running” one cannot infer “therefore this man except Socrates is running” (although this unacceptable inference is uninterpretable rather than invalid).

Supposition theory was an attempt to develop a unified treatment of a great number of semantical and logical topics that are still of interest, although now for the most part they are treated in separate inquiries. It is therefore especially intriguing and difficult to discover just what that unifying notion—supposition—amounted to. One broad description that plainly holds good for most of the divisions of supposition is that they are syntax-dependent referential functions of a term’s signification. Any description in terms of syntax and semantics will, however, fail to cover all the divisions of supposition and will miss what is distinctive in it. In the case of improper supposition, for example, while the circumstances under which a term is used clearly do determine whether or not it is being used figuratively, it will not do to limit those circumstances to the syntactic context of the term’s use.

Again, supposition theorists frequently remarked on the fact that the supposition apparently determined by the syntax often differs from the supposition intended by the framer of the statement. The man who visits his friend’s garden and says “this plant grows in my garden” says what is false “with respect to discourse” (*de virtute sermonis*), but the circumstances of his utterance show that he intends to use the word *plant* as in the statement “such a plant grows in my garden.” The correct analysis of the supposition of “plant” takes it to have the supposition determined for it *not* by that syntactic context but by the clearly discernible “intention of the framer” (*intentio ponentis*). Finally, among later supposition theorists it was a matter of controversy whether terms occurring in statements written in a closed book had any supposition at all, and those few who held that they did then have supposition seem to have been motivated by a misguided concern that otherwise certain true statements in a closed book, such as “God exists,” might cease to be true while the book remained unread.

The consensus of the modernists seems to have been that a term had one or another kind of supposition only on an occasion of its actually being used in referring (or understood to have been used in referring) to some entity or entities, the particular kind of supposition being determined by a number of the circumstances of the occasion and its syntactic context being the most important but not in itself the decisive circumstance.

## SYNCATEGOREMATA

Within the *logica moderna* the investigation of syncategorematic words complemented the investigation of the categorematic words under the doctrine of “the properties of terms.” Something closely resembling the modernists’ notion of syncategoremata seems to have been operative in Aristotle and the Stoics, but the medievals evidently acquired the technical term and the rudiments of the notion directly from Priscian (fl. 500), who had reported that “according to the dialecticians [not identified], there are two [principal (?)] parts to a sentence—the noun and the verb—since they alone and of themselves make a sentence; but they called the other parts *syncategoremata*—that is, consignificants” (*Institutio de Arte Grammatica*, M. Hertz, ed., in H. Kiel, *Grammatici Latini*, Leipzig, 1855, Vol. II, p. 54).

Interest in the syncategoremata as such began in connection with the twelfth-century interest in fallacies of ambiguity as it became plain that the crucial ambiguity was often located elsewhere than in subjects and predicates. The grammatical basis of distinction provided by Priscian seems to have been adopted at first and occasionally even narrowed so that only the “indeclinables”—prepositions and conjunctions—were considered to be syncategoremata (see the anonymous late twelfth-century *Fallacie Parvipontane*, edited by L. M. de Rijk, in *Logica Modernorum*, Vol. I, p. 559; also see Abelard, *Dialectica*, edited by L. M. de Rijk, pp. 118–121). The notion of syncategoremata that became important in the *logica moderna*, however, was not founded on a strictly grammatical distinction. Abelard’s treatment of “alone” in “a man alone is capable of laughter” (*Logica “Ingredientibus,”* edited by B. Geyer, p. 483) prefigured the pattern that was to be followed by the modernists. He pointed out that if “alone” is taken to be part of the subject, the statement is about a man who happens to be by himself, while if “it is attached to *the predication* it denies the capacity for laughter to all non-men, as if to say: ‘a man is capable of laughter in such a way that nothing else is capable of laughter.’”

When the investigation of syncategorematic words appeared as a separate inquiry in the treatises on syncategoremata by William of Sherwood and Peter of Spain (first half of the thirteenth century), the distinguishing characteristic of syncategorematic words was the fact that they had some effect on the relation between the categorematic words—that is, the predication or the “composition”—or on the relation between two predications or compositions. Thus Sherwood’s inventory of syncategoremata, most of which became standard, included the

verbs “begins” and “ceases” and the noun “nothing” as well as grammarians’ syncategoremata such as “every,” “both,” “except,” “alone,” “is,” “not,” “necessarily,” “if,” “unless,” “and,” and “or.” The standard syncategoremata, then, cannot be completely described as a selection of logical operators, although they plainly included such a selection. Nor was the investigation of them aimed primarily at uncovering their strictly formal properties. Many of the rules put forward in connection with one or another syncategorema were rules of inference—such as “When there are two distributions over one and the same part of a locution, the first immobilizes the second” (Sherwood)—but just as many were semantic rules—such as “The sign ‘every’ or ‘all’ requires that there be at least three appellata [for the term to which the sign is attached]” (Sherwood)—and there seems to have been no clear distinction drawn between the two sorts of rules.

The modernists’ treatises on syncategoremata presupposed the doctrine of the properties of terms, as is shown by the rules given just above, and much of their discussion of the function of such words is in terms of the various modifications of supposition produced within the scope of one or another syncategorema. (The problem of determining the scope of syncategoremata, especially in contexts including more than one, was particularly important in these investigations.) In this way the syntactic and semantic questions about syncategoremata were essentially interconnected. Sherwood at least among the older modernists was sometimes concerned to discuss the *signification* of syncategoremata—“every,” for example, was said by him to signify universality—but that seems to have been a feature of his unusual doctrine that in order to be significant a word had to signify some *form*. Most writers on syncategoremata took up Priscian’s really unjustified translation of the Greek *syncategoremata* as “consignificants” and used it as the basis for their view that “strictly speaking, a syncategorema signifies nothing, but when added to another word it makes that word signify something, or makes it supposit for something or some things in some definite way, or exercises some other function having to do with a categorema” (William of Ockham, *Summa Logicae*, I, 4; cf. the remarks of John of Salisbury in *Metalogicon*, Book I, Ch. 16).

The initial impetus to the study of syncategoremata came from the twelfth-century interest in fallacies, and the investigation continued to be associated with fallacies or with sophismata throughout its development. Observations about syncategoremata were only incidental in the twelfth-century treatises on fallacies, but the novel

emphasis of the *logica moderna* on the syncategoremata themselves is evident in the development of a new sort of treatise—"On Exponibles"—in the first half of the thirteenth century. "An exponible proposition is a proposition having an obscure sense that stands in need of exposition because of some syncategorema located in it explicitly or implicitly or in some word" (*Tractatus Exponibilium*, doubtfully ascribed to Peter of Spain, in *The Summulae Logicales of Peter of Spain*, edited by J. P. Mullally, p. 104). The exponibles did not involve fallacious arguments nor were they, strictly speaking, ambiguous. They were simply subjects for analysis, an analysis that was to explicate the force of some syncategorema in some particular context. For example, "'Man *inasmuch as* [*inquantum*] he is rational is not capable of braying'—that is, [1] no man is capable of braying and [2] every man is rational and [3] no rational entity is capable of braying and [4] because an entity is rational it is not capable of braying" (p. 115).

SOPHISMATA. The sophismata that played an increasingly important role in investigations of syncategoremata from the thirteenth through the fifteenth centuries may be characterized as falling somewhere between fallacies of ambiguity and exponible propositions. In the independent treatises on syncategoremata prevalent in the thirteenth century, the sophismata served as the illustrations of the principles uncovered in the investigation and characteristically took the form of an assertion (the sophisma proper) followed by a proof, a counterargument, and an adjudication of the apparent paradox by an appeal to the principles. For example, "Suppose that exactly one individual of each species of animal is running. Then [*a*] every animal is running. (Proof: a man is running; a lion ... ; a goat ... ; and so on with respect to the individuals; therefore every animal is running.) But [*b*] every man is an animal; therefore every man is running. [Solution:] [*a*] is ambiguous since [because of 'every'] the word 'animal' can distribute for the remote parts (or the individuals belonging to genera)—in which case it is false, since it is then distributed for all its individuals—or for the genera of the individuals (or for the proximate parts)—in which case the minor [*b*] is plainly not accepted" (William of Sherwood, *Syncategoremata*, edited by J. R. O'Donnell, in *Medieval Studies*, Vol. III, p. 49).

The continuity of the development of the *logica moderna* was enhanced by the fact that from the twelfth century through the fifteenth century the same sophismata were treated from varying points of view, but at the same time the number and intricacy of the sophismata

were constantly increasing. As a result the modernists of the fourteenth century frequently produced treatises titled *Sophismata* in which large numbers of them were grouped according to the syncategoremata at issue in them, and the investigations that had begun in separate treatises on syncategoremata were pursued in the *Sophismata* and *Exponibilia* of the late Middle Ages.

## SPECULATIVE GRAMMAR

The notion that grammar and philosophy were intimately related was one of the most pervasive of the assumptions that determined the character of medieval thought. It is probably to be explained by the facts that grammar was one of the very few inquiries to survive antiquity intact and that the only ancient philosophy available during the early Middle Ages was Aristotle's *Categories* and *De Interpretatione*, works of a decidedly grammatical cast. The usual view of the connection between the two subjects was the one expressed most memorably by John of Salisbury—"Grammar is the cradle of all philosophy" (*Metalogicon*, Book I, Ch. 13)—and the *logica moderna*, by far the most impressive medieval contribution to semantics, is a clear example of the influence of grammar on philosophy. The influence ran the other way, however, in the development of "speculative grammar" (*grammatica speculativa*), or the doctrine of the "modes of signifying" (*modi significandi*), a movement that began somewhat later and subsided somewhat earlier than the *logica moderna*. Although there were some connections between the two movements—for instance, Roger Bacon (1214/1220–1292), one of the first of the speculative grammarians, or "modists," also contributed to the *logica moderna*—they tended to be mutually independent and to some extent theoretically opposed developments in the history of semantics.

The most important single factor in the rise of speculative grammar in the early thirteenth century was the enthusiasm for the notion of a *science*, then being rediscovered in the *Posterior Analytics* of Aristotle and in his Arabic commentators. For a time it was the aim of every study to achieve the status of an Aristotelian science, a body of necessary knowledge deductively demonstrated, and two facts seemed to stand in the way of certifying grammar as a science. For one thing, as it had been presented by Priscian and Donatus, grammar was simply a set of observations about correct constructions without any attempt at explanation of the correctness; but only knowledge "by causes" qualified as scientific. For another, even Peter Helias (fl. 1150), who in his com-

mentaries on Priscian had been the first medieval to attempt explanations of grammatical facts, had maintained that there were as many grammars as there were languages; but a unified subject matter was a prerequisite of a science.

Thus a science of grammar was not to be found in the grammatical authorities, and it seemed one never would be found as long as grammar was conceived of as something to be discerned only in the investigation of actual languages. Robert Kilwardby (d. 1279) set the stage for speculative grammar when he argued that “since a science remains the same for all men and its subject matter remains the same, the subject matter of grammar must remain the same for all men. But grammatically ordered speech or articulate utterance that can be put into a grammatical pattern is *not* the same for all men, and for that reason it will not be the subject matter of grammar [as a science]” (Commentary on Book I of Priscian’s *Ars Minor*). No science of languages was possible, but grammar might become the science of language, the *scientia sermocinalis*, if the variable external trappings were ignored and one concentrated on the conceptual underpinnings—which, as Aristotle had pointed out, *were* the same for all men. Thus Roger Bacon was led to proclaim that “with respect to its *substance* grammar is one and the same in all languages, although it does vary *accidentally*” (*Grammatica Graeca*, edited by E. Charles, p. 278), and this became the often repeated fundamental assumption of the speculative grammarians.

As it developed, speculative grammar took the form of an attempt to provide an Aristotelian ontology of language, finding analogues in the various parts of speech for matter, form, substance, process, and so on. As Siger of Courtrai (d. 1341) put it, “grammar is the *scientia sermocinalis*, which considers discourse and its properties [*passiones*] in general for the purpose of expressing principally concepts of the mind by means of interconnected discourse” (*Summa Modorum Significandi*, edited by G. Wallerand, p. 93). Siger then cited Avicenna for the Aristotelian doctrine that concepts are the same for all men because they are the result of experiencing extramental entities, which are the same for all men. Thus, the ontology that applies to the extramental entities must apply as well to the concepts derivative from them (if they adequately copy the extramental entities) and, in turn, to the discourse employed to express those concepts (if it is to be adequate for that purpose). “Therefore *modes of being*, or properties of things ..., precede a *mode of understanding* as a cause precedes an effect” (ibid.). In the same way a mode of designating

(*modus signandi*) follows a mode of understanding, “since a thing is understood and also conceived of before it is designated by means of an *utterance* [*vox*], for utterances are signs of passions, as is said in *De Interpretatione*, Ch. 1” (p. 94). When the understanding assigns a given concept to an utterance, the merely physical utterance becomes a word (*dictio*).

Up to that point the semantic theory underlying speculative grammar might fairly be described as a technical restatement of Aristotle. It was only with the introduction of its “modes of signifying” that the theory acquired its novelty and notoriety. (It was repeatedly attacked and ridiculed by logicians and grammarians of the late Middle Ages and even more strongly assailed by the Renaissance humanists.) As an utterance becomes a word by means of a mode of designating, so a word becomes one or another part of speech by means of a mode of signifying. The modes of signifying, however, are not modes of the utterance of the word but are “certain concepts of the understanding itself” (ibid.). The kind of concept in question seems to be one that links the word to some Aristotelian mode of being. Thus, the kind of concept involved in the mode of designating is the kind that supplies a *significatum* for the utterance “horse,” transforming it from a mere sound into a word, while the mode of signifying consists in the recognition that it is substance that is signified by the word *horse*. And when the understanding adds to that general mode of signifying—substance—the specific mode of signifying—quality—then *horse* has been transformed in turn from a mere word to a substantive and from a mere substantive to a noun (pp. 94–95). Along these same lines, the utterance *horse* will eventually be accounted for as a common concrete noun, and similar patterns of modes of signifying are invoked in order to account for the other parts of speech.

Aristotelian ontology was employed by the speculative grammarians not only as the link between grammatical forms and modes of extralinguistic being but also as a picture of intralinguistic relations. Thus, verbs stood at the pinnacle of the linguistic microcosm because just as the other animals are submissive to man, so the inflections of the other parts of speech in a sentence are ultimately submissive to the verb. The infinitive of a verb, however, was analogous to primary matter in substances. And just as the organisms capable of fewest adaptations are ranked lowest in the kingdom of nature, so the indeclinables, the syncategoremata, are the most inferior parts of speech (pp. [52]–[54]).

## THE RENAISSANCE AND ENLIGHTENMENT

### SEMANTICS, LOGIC, AND EPISTEMOLOGY

As the Middle Ages gave way to the Renaissance in the late fifteenth century, logic (on which semantics had been centered) first lost its medieval attainments and then subsided into inactivity until the middle of the nineteenth century. What little there was in the way of logical inquiry from about 1450 to about 1850 was carried on under the view of logic as the art (or science) of reason, the idea of *scientia sermocinalis* having been ridiculed into oblivion by the Renaissance humanists. Aside from the work of late Scholastics, such as the *Ars Logica* of John of St. Thomas (1589–1644), and an occasional deliberate attempt at revival, such as the *Logica Fundamentis Suis a Quibus Hactenus Collapsa Fuerat Restituta* (1662) of Arnold Geulincx (1624–1669), there were no further developments of the logicosemantic theories of the *logica moderna*.

Philosophers retained their interest in semantics, however, after losing interest in and even all knowledge of the kind of logic with which it had been associated. Epistemology dominated the philosophy of the Renaissance and the Enlightenment (for present purposes, roughly 1500–1800) as logic had dominated medieval philosophy, and the development of semantics during this period centered on epistemology. As a consequence, much of the development took place in the context of discussions of nonlinguistic signs, such as representative ideas, and will not be directly considered here.

Perhaps partly because logic had lost its identity as an inquiry into language, the interest of philosophers in language was more intense and diversified during this period (and especially in the eighteenth century) than at any earlier time. Some of this interest was manifested in widespread speculation about the origin of language and in projects for a universal language or a “real characteristic.” Although works on these subjects are typical of the period and often contain material of value for the history of semantics, they can be considered here only as they bear directly on a theory of meaning or philosophy of language selected for discussion.

### BACON

Francis Bacon (1561–1626) produced comparatively little that can be described as philosophy of language, but the occasional novel insights and the programmatic charac-

ter of what he did produce helped to give it a considerable influence over philosophy of language in the Enlightenment. Almost everything of his that is relevant to the history of semantics is to be found in the “Art of Elocution or Tradition” (in the *Advancement of Learning* and the *De Augmentis Scientiarum*) and the doctrine of the “Idols of the Market Place” (in the *Novum Organum* and the *De Augmentis Scientiarum*).

The first of these is plainly Bacon’s revised version of the medieval *trivium*—grammar, logic, rhetoric—although he nowhere says so. In the later Middle Ages these subjects had sometimes been designated the *artes sermocinales*, and in the *De Augmentis Scientiarum* Bacon said that the subject matter of the *ars tradendi* was *sermo*. This inquiry into “tradition”—that is, discourse or communication—had three branches, concerning “the organ,” “the method,” and “the illustration” of tradition; and most of the work of the three branches was explicitly associated with grammar, logic, and rhetoric, respectively. For present purposes the first of these three branches is much more important than the other two.

In his scheme of “Human Philosophy” the Art of Tradition occurred as “the fourth kind of Rational Knowledge” (Spedding, Ellis, and Heath, eds., 3.383–4), because reason was “as it were the soul of discourse,” according to Bacon. “Nevertheless, in treating of them reason and discourse ought to be separated, no less than soul and body” (1.651). He began his separate treatment of discourse by identifying speech and writing as the most familiar organs of discourse and stressing their connection with reason by citing with approval the traditional version of Aristotle’s doctrine: “Words are the images of cogitations, and letters are the images of words” (3.399; but cf. 3.284, 3.85–86). But his interest in less familiar organs of discourse prompted him to frame a set of general conditions for an organ of discourse: “Whatever can be broken down into differences sufficiently numerous for explicating the variety of notions (provided those differences are perceptible to sense) can become a vehicle of cogitations from one man to another” (1.651; cf. 3.399). An organ of discourse can be used to *communicate* nothing but notions, but it will contain elements that *express* not only notions but also things.

In the most familiar arrangement of organs of discourse, words (by which Bacon meant only articulate sounds [2.411–412]) are expressed by letters—that is, phonograms. Letters, in turn, may be expressed by ciphers—that is, cryptograms—and both letters and ciphers may be designated “nominal characters.” But he recognized another kind of “notes of things, which signify



things *without* the aid or intervention of words,” either “on the basis of congruity” or “arbitrarily.” As examples of the former sort he cited hieroglyphics and gestures, gestures being “transitory hieroglyphics,” the “words” for which hieroglyphics may be the “letters,” and he classified them together as “emblems”—that is, sensible images to which intellectual conceptions could be reduced by analogy (1.652–653; 649). As examples of the latter sort he cited “real characters” such as Chinese ideograms, which “have nothing emblematic in them, but are simply surds, no less than the elements of letters themselves; ... there ought to be as many of them as there are radical words” (1.653).

Despite that disadvantage, real characters could and, Bacon thought, did function as an organ of discourse beyond the limits of a single natural language just because they signified “things and notions” without the intervention of words (1.652). Although he was convinced that there were no more convenient organs of discourse than words and letters, Bacon listed the study of the notes of things among his desiderata (1.653). Acting on this suggestion, the Royal Society commissioned some of its members to look into the project of a universal real character, the eventual result being John Wilkins’s *Essay towards a Real Character and a Philosophical Language* (1668), one of many such attempts during this period.

As another part of the inquiry into organs of discourse Bacon proposed a “philosophical grammar,” and this desideratum likewise had an extensive but un-Baconian influence. Some of what he had to say about philosophical grammar was reminiscent of the medieval speculative grammar—for instance, it was to be “a kind of grammar that would carefully inquire not into the analogy of words to one another, but into the analogy between words and things or words and reason” (1.654)—and this is what seems to have caught the imagination of his many successors in the Enlightenment who produced works in philosophical or “universal” grammar. What Bacon really had in mind was probably something more nearly like the comparative philology characteristic of the nineteenth century: “But the noblest kind of grammar would, I think, result if someone well taught in many languages, learned as well as vulgar, would treat of the various properties of *languages*, showing in what respects each excels and in what respects it is deficient” (ibid.; cf. 3.230, 3.401). He did, however, go on to suggest that one might combine all the best properties uncovered in that analysis into “a very finely formed image and remarkable model of speech itself for expressing the mind’s meanings aright” (1.654).

In his sketch of a philosophical grammar Bacon emphatically disapproved of what he believed Plato had been attempting in the *Cratylus*, an inquiry into “the imposition and original etymology of names” (ibid.; cf. 3.531), but his own concern in the doctrine of the “Idols of the Market Place” closely parallels Plato’s real concern in the *Cratylus*, that is, distinguishing between correct and incorrect names. “The idols imposed on the understanding through words are of two kinds. Either they are names of things that are not (for just as there are things that lack a name because they have not been observed, so there are names that lack things, resulting from a fantastic supposition); or they are names of things that are, but confused, ill-defined, and rashly and irregularly abstracted from the things” (1.171). As an example of the first he gave “prime mover”; his example of the second kind was “humid,” which, as his discussion of it shows, is less objectionable on these grounds now than it was in seventeenth-century English.

## HOBBS

Thomas Hobbes (1588–1679) conceived of his systematic philosophy as beginning with an investigation into language and produced different versions of the investigation in *Human Nature* (1650), Chapters 5 and 13; *Leviathan* (1651), Chapters 4–7; and *Elementa Philosophiae Sectio Prima: De Corpore* (1655; English 1656), Part I, “Computatio Sive Logica.” (The latest of those versions is also in most respects the fullest and is used as the basis of the following account.)

Philosophy, Hobbes observed, depends on ratiocination, or “computation” (Molesworth edition, 1.3). In reasoning regarding particular things “we add and subtract in our silent thoughts, without the use of words” (ibid.; see 3.32); but in most instances, and certainly in philosophizing, “men owe all their true ratiocination to the right understanding of *speech*” (1.36), such ratiocination being “nothing but reckoning, that is adding and subtracting, of the consequences of *general names*” (3.30). In the second chapter of his *Logic*, devoted specifically to “names,” Hobbes produced a novel combination of several elements in the Aristotelian-Scholastic account of the semantics of names. Ratiocination of every kind depends on memory, and the intelligent use of memory requires what Hobbes called “*marks*, namely, sensible things taken at pleasure, that, by the sense of them, such thoughts may be recalled to our mind as are like those thoughts for which we took them” (1.14).

It is possible, Hobbes thought, for a man to “spend all his time partly in reasoning and partly in inventing

marks for the help of his memory, and advancing himself in learning”—that is, to devise and profitably use a private language—but if science and philosophy are to develop, there must be “certain *signs* by which what one man finds out may be manifested and made known to others” (1.14). Signs that do “signify the cogitations and motions of our mind” are “*words* so and so connected,” or what Hobbes called “*speech*, of which every part is a *name*” (1.15). The use of names as marks, he held, was logically prior to their use as signs, since “names, though standing singly by themselves, are marks ...; but they cannot be signs otherwise than by being disposed and ordered in speech” (1.15). He recognized the syntactic disposition of names in speech as necessary but not sufficient for “declaring our conceptions to others.” Speech cannot “perform that office alone without the help of many circumstances,” such as “time, place, countenance, gesture, the counsel of the speaker” (2.274). We must “consider the drift, and occasion, and contexture of the speech, as well as the words themselves” (4.23).

When names are ordered in speech so as to be signs rather than marks, “it is manifest they are *not* signs of the things themselves” but signs only of our conceptions (1.17). Hobbes seems to have been following Aristotle’s lead here, but more faithfully than most, since he went on to say, “That the sound of this word ‘stone’ should be the sign of a stone, cannot be understood in any sense but this, that he that hears it collects that he that pronounces it thinks of a stone” (1.17). Thus, even though indirectly and only in virtue of signifying that the speaker is thinking of a stone, the name “stone” ordered in speech *is* a sign of a stone. At any event, Hobbes nowhere suggested that “stone” occurring in speech was a *name* of some mental entity. On the contrary, in going on to show that “it is not at all necessary that every name should be a name of some thing,” Hobbes began by pointing out that “‘man,’ ‘tree,’ ‘stone’ are *names* of things themselves” (1.17), though they may be used as *signs* of our conceptions of men, trees, and stones and as *names* of “fictions and phantasms of things,” such as images in dreams. “Moreover, that which neither is, nor has been, nor ever shall, or ever can be, has a name, namely, ‘that which neither is, nor has been,’ &c.; or more briefly this, ‘impossible’” (1.17). For “a name is not taken in philosophy, as in grammar, for one single word, but for any number of words put together to signify one thing” (1.23), Hobbes having decided “to apply the word ‘thing’ to whatsoever we name; as if it were all one whether that thing be truly existent, or be only feigned” (1.18).

Much of Hobbes’s investigation of names was presented in the form of discussions of traditional classifications of names. His treatment of them sometimes presents the half-understood remnants of complex medieval theories—for instance, his treatment of names of first and second intention (1.20–21)—but there are occasional interesting novelties as well. In his discussion of common and proper names he put forward his strict nominalism: “this word ‘universal’ is never the name of any thing existent in nature, nor of any idea or phantasm formed in the mind, but always the name of some word or name” (1.20); at another point he remarked that the univocal-equivocal distinction “belongs not so much to names as to those that use names” (1.23); and he based the distinction between simple and compound names not on appearances but on considerations of analyzability, so that in the context of a discussion of man “body” is a simple name while “man” is a “more compounded name,” being equivalent to “animated rational body” (1.23–24).

Hobbes encountered important difficulties in his discussion of names of “certain and determined” and of “uncertain and undetermined” signification (1.21–23), which is evidently a badly distorted remnant of supposition theory. In the course of that discussion Hobbes was led to claim, for example, that *particular* names—such as “some man”—“are of *uncertain* signification, because the hearer knows not what thing it is the speaker would have him conceive” (1.22), as if the “uncertainty” in, say, “some man will marry my daughter” were the sort that could always be resolved by asking the speaker “*which* man?” Even worse confusion resulted from his attempt to show that such quantifiers as “every” and “some” were unnecessary for purposes of reasoning. Such words, he maintained, “which denote universality and particularity, are not names; so that ‘every man’ and ‘that man which the hearer conceives in his mind’ are all one; and ‘some man’ and ‘that man which the speaker thought of’ signify the same. From whence it is evident, that the use of signs of this kind, is not for a man’s ... getting of knowledge by his own private meditation (for every man has his own thoughts sufficiently determined without such helps as these) but ... for the teaching and signifying our conceptions to others” (1.22).

In his treatment of propositions Hobbes sometimes spoke as if only such propositions as “Cicero is Tully” were true—for instance, “that proposition only is true in which are copulated two names of one and the same thing” (1.57)—but usually his description of a true proposition was more moderately and more accurately expressed along such lines as these: “A *true* proposition is

that, whose predicate contains, or comprehends its subject, or whose predicate is a name of every thing, of which the subject is a name" (1.35; cf. 4.23–24). He produced a detailed analysis of falsity as reducible to combinations of names of different sorts of entities (1.57–62). His truth theory was, however, quite radical in other respects. The "first truths," he claimed, "were arbitrarily made by those that first of all imposed names upon things, or received them from the imposition of others. For it is true [for example] that *man is a living creature*, but it is for this reason, that it pleased men to impose both those names on the same thing" (1.36). This suggests an identification of the proposition with a particular sequence of words, but Hobbes elsewhere gave the impression of having been on the point of drawing a clear distinction between propositions and the vehicles of their expression—for instance, "every proposition may be, and uses to be, pronounced and written in many forms.... And therefore, whensoever they [students of philosophy] meet with any obscure proposition, they ought to reduce it to its most simple and categorical form" (1.39).

Hobbes rejected the analysis of contingent categorical propositions into their corresponding hypothetical forms, pointing out that while this analysis was allowable for *necessary* categoricals, "in contingent propositions, though this be true, 'every crow is black,' yet this, 'if any thing be a crow, the same is black' [i.e., '(x)(Cx ⊃ Bx)'], is false" (ibid.). In several places Hobbes discussed the various uses of speech—for example, *Human Nature*, Chapter 13—and at one point argued against the notion that a promise simply by its form of words creates an obligation (2.18–20).

### THE PORT-ROYAL LOGIC

René Descartes (1596–1650) had very little to say about language, but Antoine Arnauld (1612–1694) took an avowedly Cartesian approach to semantic questions in the *Port-Royal Grammar* (*Grammaire générale et raisonnée*, with Claude Lancelot, 1660) and the *Port-Royal Logic* (*Logique ou l'art de penser*, with Nicole, 1662). The latter book had a tremendous influence; it marks, better than any other, the abandonment of the medieval doctrine of an essential connection between logic and semantics. Disdain for medieval theories was emphatically expressed in it—"No one, thank God, is interested in ... second intentions" (*Premier Discours*)—but at several points the theories were still employed (for instance, in 1.2 and 2.10) and elsewhere in the book they were supplanted by innovations that sometimes obscured what had been clear in the *logica moderna*.

In words reminiscent of Hobbes's on this point, Arnauld remarked that if logic considered only an individual's reflections on his ideas, the investigation of language would form no part of it. But we must use "exterior signs" for communication, "and since this custom is so strong that even when we think by ourselves things are presented to our mind only together with the words with which we are accustomed to adorn them in speaking to others, it is necessary in logic to consider the ideas joined to words and the words joined to ideas" (introduction; cf. Descartes, *Principles*, Part I, Principle 74). Arnauld of course argued (1.1) against Hobbes's anti-Cartesian suggestion that reasoning might be "nothing more than the uniting and stringing together of names or designations by the word 'is,'" so that all we could ever conclude is "whether or not there is a convention (arbitrarily made about their meanings) according to which we join these names together" (*Objections to Descartes's Meditations*, 3.4).

Signs and signification were frequently discussed in the *Port-Royal Logic*, sometimes with interesting results; the most fundamental questions were, however, treated with the kind of inattention to detail that came to characterize most of the many semantic theories of the Enlightenment. "The sign," said Arnauld, "comprises two ideas—one of the thing that represents, the other of the thing represented—and its nature consists in exciting the second by means of the first" (1.4). In the case of words, the "thing represented" was identified as a "thought" or an "idea." Even proper names were defined as those "that serve to mark ... the ideas that represent only one single thing," and "general words" were said to be those "that are joined to universal and general ideas" (1.6). The doctrine is so far consistent and recognizably Cartesian, even if crude. But it is complicated, no doubt inadvertently, by many suggestions of a different sort of signification for words. Thus, on a single page Arnauld began by calling words "sounds that are intended to signify *ideas*," went on to speak of "things and modes" as "the *objects* of our thoughts," and ended by defining names as "the words intended to signify both *things and modes*" (2.1); and he nowhere provided an account that might justify this extended use of "signify." He may have been assuming a transitivity of signification—words signifying ideas representing things—but John Locke was the first to attempt to spell out such a theory.

Arnauld warned against the "great equivocation in the word 'arbitrary' when we say that the signification of words is arbitrary," pointing out that while "it is purely arbitrary to join one idea to one sound rather than to

another,” nevertheless the ideas, “at least those that are clear and distinct,” are not arbitrary. The result of correct reasoning is “a solid, effective judgment regarding the nature of things based on the consideration of the ideas of them a man has in mind, which ideas it has pleased men to mark by means of certain names” (1.1). But, on the other hand, one of his reasons for rejecting Aristotle’s categories was that they were “arbitrary names that form no clear and distinct idea in the mind” (1.3).

Arnauld also explicitly rejected Aristotle’s definition of a verb, putting in its place one that not only captured the essence that Aristotle missed but was also much simpler: “a word that signifies affirmation” (2.2). Evidently he did not mean that it signified the idea of affirmation (as would the noun “affirmation”), but he did not work out the definition in a way that tied it to the rest of his signification theory.

Like so many other philosophers of the period, Arnauld believed that “the best means of avoiding the confusion of words to be found in ordinary languages is to make a new language and new words that would be attached only to the ideas we want them to represent.” He differed from most, however, in suggesting that this be accomplished simply by a conscientious, systematic use of precise nominal definitions attached to already extant vocables of ordinary languages (1.12).

One of the more interesting notions in Arnauld’s doctrine of signification was introduced in his observation that “it often happens that besides the principal idea (which is regarded as its *proper* signification) a word excites several other ideas that may be called *accessory*.” Sometimes the accessory ideas are attached to the words “as the result of common usage,” as in “you lied,” the *proper* signification of which is the idea that you knew the contrary of what you said, the ideas of contempt and outrage being *accessory* (1.14). In some respects this is reminiscent of Augustine’s doctrine in *Principia Dialecticae*, especially when Arnauld uses it to argue (against Cicero) that certain words may, in virtue of their accessory ideas, be described as unchaste (1.14). Accessory ideas may also be attached for the purpose of a single use of a word, and on that basis Arnauld attempted an explanation of the varying signification of the demonstrative pronoun “this,” here as elsewhere in the book applying his semantic doctrines to the elucidation of the formula of transubstantiation—“this is my body” (1.15).

Arnauld’s notion of accessory ideas might have been (but was not) used to advantage in his discussion of problems of identity of reference, where he argued that when the mind frames the proposition “that Rome, which was

of brick before the time of Augustus, was of marble when he died, the word ‘Rome,’ which appears as only one subject, nevertheless marks two subjects that are really distinct but reunited under a confused idea of Rome that prevents the mind from perceiving the distinction of subjects” (2.12). The suggestion is that the proposition should be rejected by anyone having a clear and distinct idea of Rome, which is preposterous. Even if the *proper* signification of “Rome” was taken to be only the idea of buildings, surely such *accessory* ideas as location, population, and institutions could have been invoked to warrant the continuing use of the single proper name.

Like Hobbes, Arnauld recognized complex terms expressed in a single word, but instead of Hobbes’s criterion of analyzability Arnauld employed the notion of accessory ideas attaching to the word under certain circumstances. Thus, the term *king*, which is simple “*in expression*,” was “a term complex *in sense*” when uttered in seventeenth-century France, “because in pronouncing the word ‘King’ we not only have in mind the general idea corresponding to that word; we also mentally join to it the idea of Louis XIV, who is now King of France. There is an infinity of terms in ordinary human discourse that are complex in this respect” (1.8).

Arnauld’s analysis of the semantics of sentences clearly illustrates the importance of the loss of supposition theory. In one badly confused but typical passage he claimed that “when one says that men are animals, the word ‘animal’ no longer signifies all animals, but only the animals that are men” (2.17). Not only does this transform predication into identity, it also violates his own doctrine of signification. Again, in discussing “some man is just” Arnauld maintained that “just” there “signifies only the justice that is in some man,” the result being that “some man is identified with some just [thing]” (2.18).

A complete chapter of the *Port-Royal Logic* is devoted to the discussion of propositions such as “this is Alexander” (pointing to his portrait), which he described as “expressions in which one uses the name of the thing to mark the sign,” seldom if ever causing any difficulty in actual use, and propositions such as Joseph’s explanation of Pharaoh’s dream—“the seven full sheaves are seven full years of plenty”—“expressions in which, the sign being marked by its own name or by a pronoun, one affirms of it the signified thing.” One result of this novel approach to metaphor is his formulation of a rule governing the appropriateness of the second sort of proposition: “the mind of those to whom one speaks must already regard the sign as a sign and be concerned to know of what it is a sign” (2.14).

“COMPREHENSION” AND “EXTENSION.” Certainly the most influential semantic doctrine of the *Port-Royal Logic* was Arnauld’s introduction of a distinction between the “comprehension” and the “extension” of a term. (In the nineteenth century Sir William Hamilton renamed the former “intension.”) The principle of such a distinction had been employed in the medieval distinction between simple and personal supposition, and even the distinction itself had occasionally been anticipated in one form or another (for instance, by Cajetan), but Arnauld’s introduction of it seems to have been original and certainly was the first instance of a systematic use of it.

It is difficult, however, to say exactly what Arnauld intended by the distinction, for its exposition is obscured by his generally confused account of signification. He first advanced the distinction as one pertaining to “universal ideas” (or terms). “I call the *comprehension* of the idea the attributes it comprises in itself that cannot be removed from it without destroying it, as the comprehension of the idea of the triangle comprises extension, figure, three lines, three angles, the equality of those three angles to two right angles, etc. I call the *extension* of the idea the subjects with which that idea agrees, ... as the idea of triangle is extended to all the various species of triangle.” And Arnauld went on to say that the idea could be restricted in its extension by “applying it to only some of the subjects with which it agrees, without thereby destroying it,” for example, by attaching to it “an indistinct and indeterminate idea of a part, as when I say ‘some triangle’” (1.6). If, however, the extension consists of species and not of the individuals, which is what Arnauld maintained, then such a device for restricting extension is always to be read as “some (species of) triangle,” which produces an absurdity. Because of his theory of signification there would be theoretical difficulties for Arnauld in simply identifying the term’s extension with the individuals in question, but for the most part he seems to have had that identification in mind rather than the one he laid down.

Individual terms, too, were said to have comprehension and extension. In the phrase “Julius Caesar, the greatest commander the world has ever seen,” the comprehension of that individual term is “explicated” in one of countless possible ways. But the extension of an individual term cannot be restricted, Arnauld maintained, and thus every singular proposition is universal (1.8, 2.3).

## LOCKE

In the third book of his *Essay Human Understanding* (1690) John Locke (1632–1704) produced the first mod-

ern treatise devoted specifically to philosophy of language. No work had a greater influence over the development of semantics during the Enlightenment than did Book III of this work, “Of Words”; yet its semantic theories were neither novel in principle nor clearly and thoroughly developed. To go no further back, many of its principles had been anticipated in Kenelm Digby’s *Two Treatises* (1664), in Richard Burthogge’s *Organum Vetus et Novum* (1678), and in Hobbes’s works. Of course Locke’s “Of Words” acquired importance simply by being a part of the enormously influential *Essay*, but the source of its special influence lay in the fact that Locke had expressly connected semantic inquiry with theory of knowledge. He had set out to investigate “our knowledge,” and along the way he found himself unexpectedly compelled to investigate “the force and manner of signification” of words (3.9.21), having discovered that “there is so close a connexion between ideas and words ... that it is impossible to speak clearly and distinctly of our knowledge, which all consists in propositions, without considering, first, the nature, use, and signification of Language” (2.33.19). The new epistemological orientation of semantic inquiries, apparent even in the logic books of the period, was first explicitly established in Locke’s *Essay*.

Locke evidently thought of the material of Book III as serving two purposes in his philosophy. On the one hand, he characterized his new “way of ideas” as nothing more than “the old way of speaking intelligibly” (third letter to Stillingfleet), which he reduced to a few commonsensical maxims for the avoidance of “jargon,” very much in the spirit of Bacon’s treatment of the “Idols of the Market Place.” The semantic theory in Book III was developed in part as a support for these “remedies of the ... imperfections and abuses of words” (3.11), and Locke’s preoccupation with that practical aim may help to explain some of the imprecision and inconsistency in his theoretical statements. He did, however, clearly recognize a more strictly theoretical purpose in the semantic inquiries of Book III, one which he summarized in his description of the third branch of science—“*Σημειωτική*, or *the doctrine of signs*” (4.21.4), the consideration of ideas as the signs of things and of words as the signs of ideas.

Locke’s account of words as the signs of ideas shows little of the sensitivity to the complexities of language that had characterized the work of many of his predecessors, including Hobbes. Except for one very short, cryptic chapter on “particles” (by which he evidently meant syntagorematic words but perhaps also verbs), the seman-

tics of words in Book III is exclusively a semantics of “names”—names of “simple ideas,” of “mixed modes,” and of “natural substances”—with no suggestion that anything has been left out of consideration (3.4.1).

The development of his fundamental thesis regarding the signification of these names begins with his observing that “words being voluntary signs, they cannot be voluntary signs imposed by [a man] ... on things he knows not.” Now what a man knows is in his mind, but all that is in a man’s mind is his own ideas. Therefore, “words, in their primary or immediate signification, stand for nothing but *the ideas in the mind of him that uses them*” (3.2.2). This is not markedly different from the starting point of many earlier semantic theories, but Locke’s initially uncompromising development of it led to some extreme consequences. Men, he observed, “suppose their words to be marks of the ideas also of other men” or “to stand also for the reality of things.” Faithful to his fundamental thesis, Locke nevertheless insisted that “it is a perverting the use of words, and brings unavoidable obscurity and confusion into their signification, whenever we make them stand for anything but those ideas we have in our own minds” (3.2.4–5).

Thus, the basic semantic relation in Locke’s account of language is that of a word used by some speaker as a proper name for some idea in that speaker’s mind. It seems to follow from this doctrine that as long as one does use words in this (the only approved) way, one cannot misuse them; and Locke does sometimes suggest that in the early chapters of Book III (see, for instance, 3.2.3). Those chapters indeed present a classic formulation of what Wittgenstein was later to criticize as the notion of a “private language.”

Establishing words as proper names of ideas in the speaker’s mind fulfills the first of Locke’s two principal conditions “for the perfection of language” (3.1.3). The second was the devising of “general words,” which he thought men accomplished by using words “for signs of general ideas.” It was evident, Locke observed, that general words “do not signify barely one particular thing; for then they would not be general terms but proper names.” His account of the signification of general words is, however, severely damaged by the inclusion in those same passages of his declaration of a thoroughgoing nominalism: “things themselves ... are all of them particular in their existence, even those words and ideas which in their signification are general” (3.3.11–12).

Although many of his most careful theoretical statements ruled out any extension of the signification of a word beyond an idea in the speaker’s mind, Locke here

(and frequently in the later chapters of Book III) was apparently assuming that by virtue of signifying an idea, a word also (secondarily and indirectly, perhaps) signified whatever the idea signified. However, he never examined that assumption or even recognized it to be one. When he came to apply his theory to the discussion of various sorts of names, he often relaxed or ignored the strictures laid down in the general theory developed in the first three chapters. Thus, in his chapter on the “names of our ideas of substances” he found it convenient to say “By the word *gold* here, I must be understood to design [that is, designate] a particular piece of matter; v.g., the last guinea that was coined. For if it should stand here in its ordinary signification, for that complex idea which I or anyone else calls gold, i.e. for the nominal essence of gold, it would be jargon” (3.6.19).

When, on the other hand, Locke did apply his semantic theory strictly, he was likely to produce such surprising results as his doctrine that every generalization about a substance, such as “all gold is fixed,” means either “that fixedness is a part of the definition, i.e., part of the nominal essence the word gold stands for; and so this affirmation ‘all gold is fixed,’ contains nothing but the signification of the term gold. Or else it means, that fixedness, not being a part of the definition of the gold, is a property of the substance itself, in which case it is plain that the word gold stands in the place of a substance.... In which way of substitution it has so confused and uncertain a signification that, though this proposition—‘gold is fixed’—be in that sense an affirmation of something real, yet it is a truth will always fail us in its particular application [since we know only our idea of gold and not ‘the real essence’ of gold], and so is of no real use or certainty” (3.6.50; compare his interesting treatment of “trifling propositions” in 4.8).

Locke’s strictly subjectivist, nominalist theory of signification in the opening chapters of Book III, which gave him so much trouble in its application, may represent nothing more than his overzealous attempt to state precisely such characteristically commonsensical observations as can be found in his *Conduct of the Understanding*, Section 29, where he advised “those who would conduct their understanding right, not to take any term ... to stand for anything, till they have an *idea* of it. A word may be ... used as if it stood for some real being; but yet if he that reads cannot frame any distinct idea of that being, it is certain to him a mere empty sound without a meaning.”

(Locke’s influence is frequently discussed in the remainder of this entry. See, for instance, the sections on

Gottfried Wilhelm Leibniz, George Berkeley, and Étienne Bonnot de Condillac and on universal grammar.)

## LEIBNIZ

Gottfried Wilhelm Leibniz (1646–1716) developed some of his views on language specifically as criticisms of Locke in his *Nouveaux Essais sur l'entendement* (finished after 1709; first published 1765). One example of this *ad hoc* development is his rejection of Locke's account of "general words" as no more than devices for avoiding the proliferation of proper names. Leibniz argued that they were necessary ingredients in the "essential constitution" of languages and went so far as to claim, in an exact reversal of Locke's position, that "it is certain that all *proper* or individual names were originally *appellative* or general" (3.1.3; see 3.3.5). Even in the *Nouveaux Essais*, however, most of Leibniz's views on language can be traced to considerations that lie at the center of his own philosophy.

**UNIVERSAL CHARACTERISTIC.** Perhaps the most important of the central considerations of Leibniz's philosophy is his lifelong preoccupation with the idea of a "universal characteristic," which cannot be examined here except as it bears directly on his philosophy of language. Leibniz's earlier doctrine of the characteristic (c. 1679) was "that a kind of alphabet of human thoughts can be worked out and that *everything can be discovered and judged by a comparison of the letters of this alphabet and an analysis of the words made from them*" (Gerhardt edition 7.185). Descartes, by contrast, had maintained that such a language (he never knew of Leibniz's scheme, of course) depended on the prior establishment of "the true philosophy" (letter to Marin Mersenne [1629], in Adam and Tannery edition 1.76).

Leibniz's initial response was that while the establishment of the characteristic "does depend on the true philosophy, it does not depend on its completion"; for as long as we have the true "alphabet of human thought" to begin with, we can complete the true philosophy simply by correctly manipulating the characteristic (Couturat edition, pp. 27–28). (The many artificial languages projected during the Enlightenment may be classified as "Cartesian" or "Leibnizian," depending on whether they were put forward solely as devices for recording and communicating knowledge or also as heuristic devices. It is the Leibnizian rather than the Cartesian projects that bear a significant resemblance in principle to the formalized languages for logic developed after the middle of the nineteenth century.) Writing some years later (1697) and in a context where the issue between his own and

Descartes's views was not explicit, Leibniz did nevertheless acknowledge that "genuinely real, philosophic characters must correspond to the analysis of thoughts. It is true that such characters would presuppose the true philosophy, and it is only now [when he believed himself to have discovered the principles of the true philosophy] that I should dare to undertake the construction of them" (Gerhardt edition 3.216).

By a "real" characteristic Leibniz meant a symbolism that was in some important respect naturally (rather than conventionally) associated with what it symbolized. Although a thoroughly real characteristic could be developed only in an artificial language, Leibniz observed that natural languages were in certain respects real characteristics. It was on the basis of that observation that he became the first major philosopher after Epicurus to suggest an appeal to ordinary language as a philosophical technique. His general attitude is expressed in the *Nouveaux Essais*: "I truly think that languages are the best mirror of the human mind and that an exact analysis of the signification of words would make known the operations of the understanding better than would anything else" (3.7.6). Part of what he meant by "exact analysis" closely resembled Plato's use of etymology in the *Cratylus*.

In his preface to a 1670 edition of Nizolius (in which he has a great deal to say about language) Leibniz argued that "the good grammarian and the philosopher too can, so to speak, deduce the use of a word from its origin by means of an unbroken sorites of metaphors" (Gerhardt edition 4.140). But he also viewed ordinary language in its unanalyzed state as having a special philosophic value: "Whatever cannot be explicated by means of popular terms (unless like many kinds of colors, odors, and tastes, it consists in immediate sensation) is nothing, and should be kept away from philosophy as if by a kind of purifying incantation" (4.143). Not every ordinary language was equally valuable as a touchstone for philosophy: "No language in Europe is better suited than German for this certifying trial and examination of philosophical doctrines by means of a living language, for German is richest and most nearly complete in real characters [*in realibus*], to the envy of all other languages.... On the other hand, the German language is easily the least well suited for expressing fabrications [*commentitia*]" (4.144; cf. Duclos edition 6.2.10 ff.). Leibniz's praise of German for its high proportion of real characters was very likely based simply on the fact that it contains words of Germanic origin where English and the Romance languages are likely to have words of Greek and Roman origin—for instance,

*Unabhängigkeit* and “independence”—a feature of the German language which no doubt does provide its native speakers with comparatively easy access to many abstract notions.

Leibniz also recognized a more pervasive kind of “realness” in natural languages that might be called *syntactic*, in contrast with the historically more familiar kind just discussed. It constituted the essential ingredient in his doctrine of “expression” and thus formed part of his metaphysics (monads *express* the universe) as well as of his philosophy of language. In *What Is an Idea?* (1678) he offered this account:

That is said to *express* a given thing in which there are relations [*habitudines*] that correspond to the relations belonging to the thing expressed. But these expressions are of different kinds—e.g., the model of a machine expresses the machine itself; the projective delineation of a thing in a plane expresses the solid; discourse expresses thoughts and truths; characters express numbers; an algebraic equation expresses the circle (or some other figure)—and what is common to these expressions is the fact that we can pass from the mere consideration of the expressed relations to a knowledge of the corresponding properties of the thing being expressed.

Leibniz drew the conclusion that “it is clearly not necessary that that which expresses be similar to that which is expressed as long as a certain analogy of [internal] relations is preserved” (Gerhardt edition 7.263–264). What he was proposing, however, was clearly a novel approach to resemblance as a basis for semantic relations, suggesting for the first time that in complex signs the “realness” of the symbolism may consist in the resemblance between the *schemata* of the expression and of what is expressed rather than in a resemblance between the *elements* of those two schemata. This was brought out most clearly in his *Dialogue* (1677)—for example, in the observation that “even if the *characters* are arbitrary, still *the use and interconnection* of them has something that is not arbitrary—viz. a certain proportion between the characters and the things, and the relations among different characters expressing the same things. This proportion or relation is the foundation of truth” (7.192).

In describing this schematic resemblance as the foundation of truth, Leibniz stated the principal thesis in his novel doctrine of propositions as extralinguistic, extramental schemata. Although such a notion of propositions had been hinted at by Hobbes, Leibniz was evi-

dently the first to make it explicit; and, as it happened, he developed his doctrine in conscious opposition to Hobbes’s view of truth as dependent on words and hence arbitrary. It had been standard philosophical usage from the beginning of the Middle Ages to use the word *proposition* for whatever was either true or false, and the principal refinement of this usage before Leibniz had been the medieval distinction of “mental” propositions from propositions spoken or written. Leibniz’s first objection against what he called Hobbes’s “super-nominalism” might be interpreted as going no further than that, as in his observation that “truths remain the same even if the notations vary” (*Preface to Nizolius* [1670], in Gerhardt edition 4.158). He subsequently recognized, however, that those “truths” could not be identified with true propositions that had been, were, or would be actually in someone’s mind—for instance, in the *Dialogue* of 1677: “A. ... Do you think that all propositions are thought? / B. I do not. / A. You see, therefore, that truth does belong to propositions or thoughts, but to *possible* [propositions or thoughts], so that this at least is certain, viz. that *if* anyone should think in this [or a contrary] way, his thought would be true [or false]” (7.190).

Once Leibniz had distinguished propositions from actual thoughts and from combinations of words, he was in a position to reject the traditional account of truth as “the conjunction or separation of signs according as the things themselves agree or disagree among themselves,” in which account “by ‘the conjunction or separation of signs’ one must understand what is otherwise called a proposition.” Leibniz’s attack on this tradition contrasted its technical terminology with ordinary usage in order to show that it concealed rather than resolved problems:

An epithet—e.g., “the wise man”—does not make a proposition, and yet it is a *conjunction* of two terms. Negation, moreover, is something different from *separation*, for saying “the man” and after an interval pronouncing “wise” is not to deny. Finally, *agreement* [or *disagreement*] is not, strictly speaking, what one expresses by means of a proposition; two eggs have agreement and two enemies have disagreement. The manner of agreeing [or disagreeing] at issue here is quite extraordinary [*toute particulière*]. Thus I think this definition completely fails to explicate the point at issue. (*Nouveaux Essais* 4.5.2)

And Leibniz went on from this criticism of the traditional doctrine of propositions to present once again his own



view of them as entities distinguishable both from words and from actual ideas (*Nouveaux Essais* 4.5.2).

**“LEIBNIZ’S LAW.”** Leibniz’s famous principle of substitutivity, known in recent literature as Leibniz’s Law, was frequently used as a starting point by twentieth-century writers on semantics. Leibniz employed the principle as part of the primitive basis of his logical calculus and put forward several versions of it in papers written from 1679 through the early 1690s. The various versions may be accurately synthesized as follows: Those entities are the same, one of which may be everywhere substituted for the other, preserving the truth(-value) (see 7.219; 7.228; 7.236).

Although Leibniz did not identify the entities in question and sometimes discussed the principle as if it applied, for example, to geometrical figures, the context generally makes it plain that its principal intended application was to terms in propositions actually expressed in some notation. His discussion of the principle in the papers in which he applied it took no account of contexts in which the principle does not apply, but at least one passage in his later writings shows that he had by then recognized that cases of what the medievals had called material supposition did not fall under the principle. “Indeed, one sometimes speaks of words *materially* without being able in that context [*cet endroit-là*] to substitute in place of a word its signification, or its relation to ideas or to things. This occurs not only when one speaks as a grammarian but also when one speaks as a lexicographer, in giving the explication of a name” (*Nouveaux Essais* 3.2.6). Recent criticism of Leibniz’s Law has often begun with the complaint that he failed to notice just such exceptions.

**Berkeley.** Locke had argued that a word was significant solely in virtue of standing for an idea in the mind of the user of the word. When George Berkeley (1685–1753) began philosophizing, he accepted that doctrine as axiomatic. In several early entries in his private *Philosophical Commentaries* (1707–1708) he presented it as part of the basis of his otherwise anti-Lockean position—for instance, “All significant words stand for Ideas” (Luce and Jessop edition 1.45; see 1.39, 1.43, 1.53). Even before ending the *Commentaries*, however, Berkeley had rejected Locke’s semantics too and had begun to replace it with a doctrine of great importance in the development of his own philosophy and in the history of semantics.

The actual turning point was apparently reached in his discovery that some words that should have been paradigm cases for Locke’s semantics had no precisely corre-

spondent ideas. “Qu: How can all words be said to stand for ideas? The word Blue stands for a Colour without any extension, or abstract from extension. But we have not an idea of Colour without extension; we cannot imagine Colour without extension” (1.62). In this passage Berkeley questioned for the first time not only Locke’s semantics but also (indirectly) his doctrine of abstract ideas. He very soon saw that the connection between the two was essential. Given Locke’s semantics, together with the facts that a general word was significant and that no concrete particular idea corresponded to it, one was forced to introduce a Lockean abstract idea simply in order to give a general word something to stand for. (As Berkeley pointed out [2.36], Locke had virtually admitted as much in the *Essay* [3.6.39].) Berkeley’s alternative account in the Introduction to *The Principles of Human Knowledge* (1710), Section 11, was that “a word becomes general by being made the sign, not of an abstract general idea but, of several particular ideas, any one of which it indifferently suggests to the mind” (2.31; see 2.127). Berkeley’s account thus involved abandoning Locke’s semantic principle that there be a single idea to serve as the name-bearer for each significant word.

In the history of semantics, however, as in the history of philosophy in general, Berkeley’s rejection of abstract ideas is more important than his alternative account of the signification of general words. The rejection was based not only on the well-known exposition of the internal inconsistency—as in *Principles*, Introduction, Section 13 (2.32–33)—but also on his many and varied attacks on their semantic foundation. Since Locke’s commitment to the view that each word had to stand for one idea in order to be significant was what had compelled him to introduce abstract ideas, Berkeley set out to show, by means of various sorts of counterinstances, “that words may be significant, although they do not stand for ideas. The contrary whereof having been presumed seems to have produced the doctrine of abstract ideas” (3.292–293; see 1.70).

He seems to have found at least four sorts of counterinstances, the first and most obvious consisting of words that stand for something other than ideas. Words such as “volition,” “I,” “person,” and the “particles” (or syncategorematic words) are significant in virtue of standing for “spirits” or their activities (the particles standing for “the operations of the mind”) (1.65, 1.80, 1.81; see 3.292). But in Berkeley’s immaterialism there were no entities other than spirits and ideas for which words could stand, and so there could be no other counterinstances consisting simply of words that stood for

nonideas. He was thus led to investigate the relation “stands for” more closely than its relata. In a move reminiscent of supposition theory he attacked Locke’s account of “understanding propositions by perceiving the agreement or disagreement of the ideas marked by their terms,” claiming that when he asserted of a particular dog “Melampus is an animal” he had not two ideas but “only one naked and bare idea, viz. that particular one to which I gave the name Melampus.” Nor does “animal” in that proposition “stand for any idea at all. All that I intend to signify being only this, that the particular thing I call Melampus has a right to be called by the name animal.” But it would not do, he pointed out, to say that “animal” here stood for the same idea as did “Melampus,” since that would make the proposition a tautology (2.136–137; cf. 1.69, 8).

The principal effect of this second sort of counterinstance was to raise some serious doubt regarding the nature of the relation “stands for,” and Berkeley’s remaining counterinstances took the almost unprecedented step of suggesting that that relation was not always an essential ingredient in significance. Words that might in certain occurrences be said to stand for ideas are very often used in reasoning and in ordinary conversation as uninterpreted (but interpretable) “counters.” A word used in that way does *not* in each of its occurrences stand for an idea in the mind of the user or, for that matter, raise a corresponding idea in the mind of the hearer or reader (2.37, 3.291–292, 8.25, 8.27).

Finally, a word sometimes occurs in a context such that one would miss rather than grasp its significance by taking it to stand for the idea to which it is customarily attached. “For example, when a Schoolman tells me *Aristotle hath said it*, all I conceive he means by it, is to dispose me to embrace his opinion with the deference and submission which custom has annexed to that name” (2.38). What is more, a word may occur in a context that precludes the possibility of taking it to stand for an idea without thereby being rendered insignificant—for example, the subject term in “the good things which God hath prepared for them that love him are such as eye hath not seen nor ear heard nor hath it entered into the heart of man to conceive.”

It was Berkeley’s view that the significance of propositions such as these last two was to be found not in the ideas the words might otherwise be said to stand for but in the purpose, or “design,” of the proposition. The design of this last example cannot be “to raise in the minds of men the abstract ideas of thing or good nor yet the particular ideas of the joys of the blessed. *The design is to*

*make them more cheerful and fervent in their duty*” (2.137 [italics added]; see 2.293, 3.292). Words, he held, “have other uses besides barely standing for and exhibiting ideas, such as raising proper emotions, producing certain dispositions or habits of mind, and directing our actions” (3.307). Thus, in his attacks on the semantic foundation of Locke’s doctrine of abstract ideas Berkeley came nearer than anyone since the Stoics to abandoning, or at least supplementing, the attempt to account for all linguistic meaning in terms of the relation between names and their bearers.

As for Locke’s semantics, Berkeley had reduced it to the unexceptionable principle of common sense that had no doubt prompted Locke’s theoretical claims, namely, we ought not to use words without knowing their meaning (1.78; 2.76). But he took Locke’s call for a new “doctrine of signs” quite seriously, summarizing his own (mostly anti-Lockean) semantic theory under that heading in *Alciphron* (1732), Dialogue VII, Section 14 (3.307). Like Bacon, Hobbes, and Locke before him, Berkeley thought of himself as providing philosophical remedies for the abuse of words, but he differed from them in making this the core of his philosophy, announcing that “the chief thing I do or pretend to do is only to remove the mist or veil of Words” (1.78; see 2.40). He set out explicitly to do just that at many points throughout his writings, but nowhere in a more concentrated form than in the introduction to the *Principles*.

In keeping with that aim Berkeley frequently urged his readers to contemplate ideas apart from words, maintaining that “if men would lay aside words in thinking ’tis impossible they should ever mistake save only in Matters of Fact” (1.84; see 2.40). He felt, therefore, that it was “absurd to use words for the recording our thoughts to ourselves: or in our private meditations” (1.62) and introduced his “Solitary Man” for the purpose of examining that pristine state of mind in a concrete example, “to see how after long experience he would know without words” (1.71; see 2.141–142). Such passages taken together suggest an anticipation of Wittgenstein’s attack on the notion of a private language, but Berkeley had second thoughts about the absurdity of the private use of words and seems to have concluded that “the Solitary Man would ... find it necessary to make use of words to record his Ideas if not in memory or meditation yet, at least, in writing without which he could scarce retain his knowledge” (1.75).

Berkeley’s ingenious linguistic analogy in his account of sense experience, the “Universal Language of Nature,” was first put forward in his *New Theory of Vision* (1709) and developed in several later works. Speaking strictly, it

belongs to his theory of knowledge rather than to his philosophy of language, but in the course of developing the analogy he often made interesting observations about language conceived in the ordinary sense (see, for instance, 1.228–233, 1.264–265).

### MAUPERTUIS AND HIS CRITICS

In the latter half of the eighteenth century philosophical interest in language was concentrated among French philosophers. Under the influence of Condillac and the British empiricists they eventually came to consider the analysis of signification their most important task. Among the earliest figures in this development was Pierre-Louis Moreau de Maupertuis (1698–1759), who first published his brief *Réflexions philosophiques sur l'origine des langues et la signification des mots* in 1748. To some extent his position resembled those taken by Berkeley and by Condillac in his first book, *Essai sur l'origine des connaissances humaines* (1746); but Maupertuis seems to have written his *Réflexions* before he knew their work. Partly because of the author's fame as a scientist, the *Réflexions* attracted considerable attention and was commented on by Baron de L'Aulne Turgot (1750), Condillac (1752), Denis Diderot (1753), Voltaire (1753), and Maine de Biran (1815), among others.

Maupertuis conceived of the question of the origin of language very much as philosophers since Descartes had been conceiving of the question of the origin of knowledge. It was intended to give rise not to speculations about prehistoric man but rather to an analysis of the hypothetical circumstances of a man with fully developed faculties who has suddenly been deprived of all his memories and of all human society. Would such an individual frame a language at all? If he did so, what would be the stages of its development? By asking and answering such questions as these within the framework of his "metaphysical experiment," Maupertuis expected to gain insight into the nature of language and its relation to the acquisition of knowledge. He began by imagining himself in the condition of the adult newborn. As soon as he had had two perceptions,

I should see that the one was not the other, and I should try to distinguish between them. And since I should have no ready-made language, I should distinguish between them by means of any marks whatever and might be satisfied with the expressions "A" and "B" as standing for the same things I now mean when I say "I see a tree," "I see a horse." Receiving new perceptions after-

wards, I could designate them all in that way.  
(*Réflexions*, Sec. 7)

It is not clear whether his saying "A" to himself in this protolinguistic context is really separable from his act of individuating the perceptual event of his seeing a tree, but Maupertuis did consider "A" and "B" as signs of his perceptions and thus presented a nearly classic case of what Wittgenstein later described as a "private language." The first development beyond those initial "signs" was recognized by Maupertuis as sufficiently radical to be described as "another language." In Section 8 he wrote:

For example, in the preceding perceptions I should recognize that each of the first two had certain characteristics that were the same in both and that I could designate those by a single sign. Thus I should change my first simple expressions "A" and "B" into these: "CD" and "CE," which would differ from the first only in that new convention, and which would correspond to the perceptions I now have when I say "I see a tree," "I see a horse."

Maupertuis's analysis proceeded in this way until he had introduced devices for discriminating kinds of perception, numbers of objects perceived, remembered and anticipated perceptions, and so on. His purpose in doing so, however, was to provide the background for a new philosophic method, which he applied most notably in his analysis of "the force of the proposition 'there is ...'" Although in saying "there is a tree" I may seem to be making a claim that goes beyond the evidence of my perceptions, once language has been reconstructed on the basis of my perceptions alone, I am in a position to see that "there is a tree" is no more than an abbreviation for "I shall see a tree every time I go to that place." This latter proposition in turn is reducible to the sequence "I was in a certain place," "I saw a tree," "I returned to that place," "I saw that same tree again," and so on (Secs. 24–28).

Eight years after writing the *Réflexions* and having meanwhile read the French translation of Berkeley's *Dialogues* (1750), Maupertuis readily admitted the similarity of his metaphysics to Berkeley's and rested his claims of independent importance on having introduced an analysis of language as the means to that end. "The point is that this philosopher [Berkeley] attacks the system of our errors only by parts. He demolishes the structure at the top, and we undermine its foundations. This is a structure quite different from that famous tower the erection of which on the Plains of Shinar was prevented by the confusion of tongues; this one is not erected *except by abusing or forgetting the meaning of words*" (*Reply to*

*Boindin*, Sec. II, in *Oeuvres*, 1756 ed.). Berkeley's own attitude was, of course, much the same; but the *Dialogues* alone among his major works fails to bring that out.

Several of Maupertuis's critics, most notably Turgot, attacked the hypothesis on which he rested his inquiry: "A solitary man such as Maupertuis imagines ... would not try to find marks with which to designate his perceptions. It is only when confronted with other people that one looks for such marks. From this there follows what is obvious in any case, that the first purpose and first step of language are to express objects and not perceptions" (*Remarques critiques*, Sec. 7). Only Maine de Biran among Maupertuis's critics defended his use of the private-language hypothesis (*Note sur les Réflexions*, Sec. V).

### CONDILLAC

Étienne Bonnot de Condillac (1715–1780) wrote his first book, *Essai sur l'origine des connaissances humaines* (1746), in an effort to do what he felt Locke might have done if he had not "realized too late" the importance for epistemology of the material in Book III of his *Essay*, "Of Words" (Locke's *Essay* 3.9.21). Locke had "treated only in his third Book what should have been the subject matter of the second" (*Essai*, edited by G. Le Roy, 1.5a). Condillac acknowledged its historical value: Locke seemed to him to have been "the first to have written on this material as a genuine philosopher. I felt, nevertheless, that it had to form a considerable portion of my own work, both because it can be viewed in a novel, more extended way and because *I am convinced that the use of signs is the principle that discloses the source [développe le germe] of all our ideas*" (1.5b; italics added). Condillac thus became the first modern philosopher to found his theory of knowledge, and consequently his entire philosophy, on considerations of signification and language, considerations that occupied him throughout his career and that shaped French philosophy for at least fifty years afterward.

Like Locke, Condillac denied that the ideas produced in sensation alone constitute a kind of knowledge, but he began his divergence from Locke in his account of the acquisition of knowledge on the basis of such ideas. "The *sole* means of acquiring knowledge is to trace our ideas back to their origin, to observe their generation, and to compare them under all possible relations. This is what I call *analysis*" (1.27a). Analysis consists in discriminating and ordering elements that are presented confusedly and simultaneously and thus requires the introduction of interrelatable signs for those elements. On these observations Condillac based his leading principle that "every

language is an analytic method and every analytic method is a language" (2.419a).

This has the look of a vicious circle. Analysis is said to be a necessary condition of knowledge, and language to be necessary for analysis; but surely knowledge is also necessary for the formation of a language. Condillac attempted to break this circle by introducing the notion of an innate language, which he called the language of action. "The elements of the language of action are born with man, and those elements are the organs given us by the author of our nature. Thus there is an innate language although there are no innate ideas. Indeed, it was necessary that the elements of some sort of language, prepared in advance, should precede our ideas, since without signs of some kind it would be impossible to analyze our thoughts" (2.396b).

In its most rudimentary form this "language" consists simply in overt reactions: "our external conformation is set to represent everything that takes place in our soul" (ibid.). Involuntary expressions of fear, pain, desire, and so on are not elements of analysis for the individual producing them, but observers of his responses can, as a result of observing the order of events making up his responses, see analyzed for them what is simply gross experience for the respondent. "Men begin to 'speak' the language of actions as soon as they feel anything, and they speak it then without having any plan of communicating their thoughts. They form the plan of speaking it in order to make themselves understood only when they notice that they have been understood" (2.397a). The usefulness of results gained by this means stimulates a natural feedback process of development on "the principle of analogy," and the language of action is made more effective by the gradual transformation of "natural" and "accidental" signs into "signs of institution," the most convenient of which are articulate sounds (1.60b–62a). The origin of language, discussed as an independent topic by many of his contemporaries and successors, is thus an essential consideration in Condillac's epistemology.

Signs of institution, including, of course, words, are themselves natural in the sense that as a language develops, they are framed on analogy with more primitive elements in that same language (and ultimately with elements of the language of action). The principle of analogy is in fact a necessary ingredient in any usable language (compare Bacon's doctrine of "emblems"). "Imagine an absolutely arbitrary language, such that analogy had determined neither the choice of words nor their various senses. That language would be an ununderstandable gibberish" (2.471a). If the principle of analogy

remained unimpaired in ordinary languages, “we would reason as nature teaches us to reason, moving effortlessly from discovery to discovery.” But every ordinary language has been impaired to some extent by the intrusion of words that have the roots of their analogy in other languages. (A similar line of reasoning had led Leibniz to praise the German language as a natural “philosophical characteristic.”) Perhaps, then, the principle of analogy can be retained in a perfectly unadulterated form only in a highly artificial language, such as algebra, which Condillac describes as “the language of mathematics.”

Since language of some sort is a necessary condition of knowledge, it is a mistake to maintain, as Locke had done, that the primary purpose of language is to communicate knowledge. “The primary purpose of language is to analyze thought. In fact we cannot exhibit the ideas that coexist in our mind successively to others except in so far as we know how to exhibit them successively to ourselves. That is to say, we know how to speak to others only in so far as we know how to speak to ourselves” (1.442a). It is a consequence of this view that the art of thinking, or logic, reduces to the art of speaking.

Although a thought is not a succession in the mind, it has a succession in discourse, where it is decomposed into as many parts as there are ideas making it up. Then we can observe what it is we do when we think, we can give an account of it, we can, consequently, learn how to conduct our reflective thought. In this way thinking becomes an art, and that art is the art of speaking. (1.403b)

Condillac’s view of the connection of thought and language was reinforced by his observations on “abstract general ideas.” “When, for example, I think about *man*, I cannot consider anything in that word except a common denomination, in which case it is perfectly plain that my idea is in some way circumscribed in that name, that it extends to nothing beyond the name, and that, consequently, it is only that name itself” (2.401b). Thus the clarity and precision of abstract ideas “depends entirely on the order in which we have produced the denominations of classes. Therefore, there is only one means of determining ideas of this sort, and that is to produce a well-made language” (ibid.). Abstract general ideas, however, are the principal ingredients of reasoning, and Condillac was even ready to say that “to speak, to reason, to produce abstract or general ideas for oneself, are at bottom one and the same thing” (2.402a). His consideration of abstract ideas, then, was one more “proof that we

reason well or badly only because our language is well or badly made” (ibid.).

All intellectual progress, on Condillac’s view, depended on and in part consisted in establishing a “well-made language,” and “a science, properly treated, is nothing other than a well-made language” (1.216a). The one perfectly well-made language so far established, he thought, was mathematics, which he examined from this point of view in his last book, *La langue des calculs* (1798). One reason why Condillac was prepared to identify a science with a well-made language is to be found in his doctrine of propositions. All that remains to be done in a science once the appropriate language has been established is the mechanical exposition of the truths proper to that science. The exposition is mechanical because “a proposition is only the unfolding of a complex idea in whole or in part,” and since a proposition “in which one and the same idea is affirmed of itself” is an identical proposition, “every truth is an identical proposition” (2.748a). An identical proposition may, however, be instructive for some persons, namely, those who observe “for the first time the relation of the terms out of which it is formed.... Thus a proposition may be identical for you and instructive for me” (2.748b). Nevertheless, “if in all the sciences we could equally trace the generation of the ideas and everywhere apprehend the true system of the things, we should see one truth give birth to all the rest, and we should find the abridged expression of all we know in this identical proposition: *the same is the same*” (2.749b).

Condillac’s influence extended not only to philosophers but also to the great chemist Antoine Lavoisier (1743–1794), who in his *Méthode de nomenclature chimique* (with Louis-Bernard Guyton de Morveau, 1787) and *Traité élémentaire de chimie* (1789) wholeheartedly adopted Condillac’s notion of a science as a well-made language. Operating under this notion, Lavoisier introduced such technical terms as “phosphoric acid” and “sulphuric acid” in a successful attempt to initiate the development of the language of modern chemistry on Condillac’s principle of analogy.

#### LAMBERT, HAMANN, AND HERDER

In the century between Leibniz and Wilhelm von Humboldt, philosophy of language in Germany was concentrated in the writings of three men: Johann Heinrich Lambert (1728–1777), Johann Georg Hamann (1730–1788), and Johann Gottfried Herder (1744–1803).

Lambert was a distinguished mathematician and the first man to follow Leibniz’s lead in his contributions to

logic, the most important of which was the earliest attempt at a calculus of relations. His work in the philosophy of language appeared in his *Neues Organon* (1764), especially Part III, “Semiotik, oder die Lehre von der Bezeichnung der Gedanken und Dinge.”

In philosophy of language, as in logic, the principal influence on Lambert was that of Leibniz, as may be seen in his preoccupation with the effect of language on thought and knowledge and with the possibility of controlling and improving that effect. Various natural languages impose various structures on our knowledge, but every natural language is fundamentally the product of prephilosophical, prescientific humankind. When we attempt to use such a language in advanced intellectual activities, we must submit our thought to the tyranny of usage (III, 1). We are thus led to seek an artificial language that from its inception could be entirely subjected to the needs of the intellect.

Lambert’s attitude toward such artificial languages differed, however, from Leibniz’s and constitutes a significant development in this line of thought. Great men, he observed, have worked at the project of a simple, perfectly regular and precise rational language, but without notable success. In any case, the adoption of such a language would be practically impossible (III, 2, 330). If we then revert to natural languages, however, we find that, strictly speaking, we cannot adopt any single one of them as a foundation for knowledge. There are, in the first place, conflicting usages even within a single natural language, some of which would have to be more or less arbitrarily ruled out; and, in the second place, any set of usages finally adopted would inevitably continue to undergo changes within the natural language of which they were a part.

Once we recognize that we do thus necessarily deviate to some extent from any given language in adapting it to intellectual purposes, it is apparent that we ought to do so consciously and under the guidance of preexamined criteria. The criteria developed and employed by Lambert were, he observed, the sort that might have served as the operative rules of a philosopher’s artificial language. In fact, he seems to have elevated Leibniz’s projected “universal characteristic” to the status of an ideal language, the principles of which are approximable to varying degrees but never fully realizable. He described his detailed examination of language as one that made a point of not distinguishing sharply between “actual and possible languages,” meaning thereby that his approach to natural language was a mixture of description and prescription in which he attempted to point out those

aspects of the actual language which were already accommodated to certain requirements of the ideal and to suggest ways in which those aspects might be enhanced and extended without introducing radical reforms that had little chance of acceptance.

The fundamental criterion employed by Lambert in his evaluation of sign systems in general and of natural languages in particular was the interchangeability of “the theory of the signs” and “the theory of the objects” signified, the degree of interchangeability marking the extent to which the signs approximated the fundamental ideal of being “scientific” (III, 23–24)—he cited musical notation as an example of a particularly close approximation. It seems evident that this fundamental criterion, which with its many corollaries pervades Lambert’s philosophy of language, constituted his adaptation of Leibniz’s doctrine of “expression.” Besides systematic general chapters on various aspects of language, Lambert’s “Semiotik” includes specific examinations of the character and function of various parts of speech and of the philosophical significance of etymological and syntactic interrelations among words.

Lambert and Hamann shared the conviction that the character of language was a topic of the greatest importance for philosophy, but they differed in almost every other important respect. Hamann’s writings are undisciplined, obscure, and strongly colored by religious mysticism. Philosophically he was a forerunner of romanticism and existentialism, consciously rejecting most of the attitudes of the Enlightenment.

To the extent to which Hamann’s philosophy exhibits a structure, it centers on his views on language, so much so that he himself called it verbalism (*Schriften*, edited by C. H. Gildemeister, 5.493–495). In almost everything he had to say about language, however, he opposed his contemporaries—Lambert (and the Leibnizian tradition) implicitly, Herder explicitly. The fundamental thesis of Hamann’s verbalism is that ordinary natural language does and should take philosophical precedence over all technical or abstract language. Occasionally he wrote as if his basis for this claim was that God had employed such language as the instrument of revelation (*Schriften*, edited by F. Roth and G. A. Wiener, 1.85–86, 1.99), but he seems to have had more generally evaluable reasons for it as well. He evidently felt that the opposition between the rationalists and the empiricists of the Enlightenment was irresolvable largely because of the reliance of both parties on introspection. The special importance of ordinary language in this connection was that it constituted a medium in which the operations of reason and experi-

ence were united and made publicly accessible. The operations of reason, indeed, consisted entirely in linguistic operations (Roth and Wiener, 6.15 and 6.25; Gildemeister, 5.515, 7.9). Philosophy, however, had traditionally adulterated what should have served as its principal source and instrument. Hamann brought this out in a characteristic attack on Immanuel Kant's abstract, technical language in the first *Critique*:

While geometry fixes and fictionalizes the ideality of its concepts of points without parts, of lines and planes conforming to ideally divided dimensions, by means of empirical signs and figures, metaphysics misuses all the word-signs and figures of speech of our empirical knowledge as mere hieroglyphs and types of ideal relations and as a result of this learned mischief transforms the straightforwardness [*Biederkeit*] of language into such a senseless, whirling, unsteady, indeterminable something (= *x*), that nothing remains but a windy murmuring, a magical shadow-play, at best ... the talisman and rosary of a transcendental superstition regarding *entia rationis*, their empty sacks and slogan. (Roth and Wiener, 7.8)

It was not only Kant's misuse of language that attracted Hamann's criticism but more especially his utter neglect of language as a topic for inquiry, which from Hamann's point of view vitiated Kant's claim to have provided a critique of reason. To point up the folly of such neglect, he tried to show that at various crucial junctures in his argument (as in the deduction of the categories) Kant had uncritically relied on certain linguistic conventions and that what he had called paralogisms and antinomies of reason really had their roots in the misuse of language.

Hamann based his doctrine of the preeminence of ordinary language not only on its value as a subject for philosophical inquiry but also on the fact that it alone among types of language was "objectively given." As such it served as the "womb" of reason and of all specialized, abstract languages designed to aid the operations of reason. Moreover, since ordinary language thus constituted the ultimate link between language and reality, all such abstract languages must be held finally accountable to it, that is, translatable into it.

As a philosopher of language, Herder is best known for his prize essay on the origin of language (1771), a topic with which this entry is generally not concerned. Herder's essay, however, occupies a position of special importance among hundreds of similar productions by

eighteenth-century philosophers, for it began the trend away from the speculative problem of origin and toward the scientifically more accessible problems of the development of language. (It was praised for that reason by several of the great linguists of the nineteenth and twentieth centuries, such as Grimm, Theodor Benfey, Edward Sapir, and Otto Jespersen.)

Ostensibly Herder was adjudicating between two rival accounts of the origin of language, but his real purpose was to dismiss the problem as senseless. The two theories at issue were those of special divine creation and of deliberate human invention of language, the former as represented in J. P. Süssmilch's work and the latter associated primarily with Jean-Jacques Rousseau's second *Discours*. Herder took reason to be the defining characteristic of man and argued in support of Hamann's position that the operation of reason and the use of language were inseparable. He then drew the obvious conclusion that if God had created what was genuinely man, He had created a language-using animal and no special divine creation of human language was conceivable. By the same token, animals correctly describable as men could not conceivably have invented language. Thus the question of how and when humans came to use language was misconceived, although the question of how primitive human languages developed was well worth considering. Hamann ridiculed Herder's argument, with justification. It seems probable, however, that the argument was intended as irony, to deflate the pretensions of the theorists rather than to refute the theories.

#### THE "IDÉOLOGUES."

Antoine Louis Claude Destutt de Tracy (1754–1836) devised the name *idéologie* for the new "science of the analysis of sensations and ideas." One section of the Second Class of the Institut National (founded in 1795) was devoted to that science in lieu of the prescientific inquiry known as metaphysics, and Destutt de Tracy and other philosophers associated with the work of that section became known as *idéologues*. For about eight years, until Napoleon Bonaparte abolished the Second Class of the Institut in the reactionary atmosphere of the First Empire, the *idéologues* were the dominant philosophical group in France. They thought of themselves as working in a field that had been opened by Locke and first thoroughly explored by Condillac, whose most original contribution was considered to be his discovery that language was as essential to the more fundamental processes of thought and analysis as it was to communication. Although part of at least Destutt de Tracy's interest in lan-

guage was directed specifically to *grammaire générale* (see below), the *idéologues* followed Condillac in considering language a topic of importance in every area of philosophical inquiry.

The *idéologues* resented being thought of as disciples of Condillac, however, and while they did, professedly, share some of his broad philosophical convictions, much of what they had to say about language (as about other topics) involved substantial revision or outright rejection of Condillac's specific doctrines. There seems to have been some tendency for the revisions to take the form of a generalization of Condillac's doctrines, notably from a concern with language to a concern with signs of all sorts. Thus, in a representative passage of his *Rapports du physique et du moral* (delivered in 1796), Pierre-Jean-Georges Cabanis (1757–1808) purported to be defending and explicating Condillac's central claim that every language was an analytic method by arguing that "one distinguishes among sensations only by attaching to them signs that represent and characterize them; one compares them only in so far as one represents by signs either their resemblances or their differences." Of course, Cabanis pointed out, taking this account as explicative of Condillac's claim required taking "language" in "the broadest sense," as meaning "the methodological system by means of which one pins down [*fixe*] one's own sensations" (*Oeuvres*, edited by Claude Lehec and Jean Cazeneuve, p. 157).

The question of the nature and epistemological function of signs (including linguistic signs) took on critical importance for the *idéologues* as a result of Destutt de Tracy's *Mémoire sur la faculté de pensée* (delivered in 1796), prompting them to set "the influence of signs on the faculty of thought" as the subject for the first essay competition sponsored by the section on the analysis of sensations and ideas. The best entries were *Des Signes envisagés relativement à leur influence sur la formation des idées* by Pierre Prévost, *Introduction à l'analyse des sciences, ou de la génération, des fondements, et des instruments de nos connaissances* by P.-F. Lancelin, and *Des Signes et de l'art de penser, considérés dans leurs rapports mutuels* by Marie-Joseph Degérando (1772–1842). Degérando's essay won the prize and was published in an expanded four-volume version in 1800. In it some of the principal issues in the philosophy of language of the seventeenth and eighteenth centuries were subjected to a final scrutiny, partly as a result of the historically apt questions provided by the *idéologues* as a guide for the essayists:

(1) Is it really the case that sensations can be transformed into ideas only by means of signs? Or, what comes to the same thing, do our first ideas depend essentially on signs? (2) Would the art of thinking be perfect if the art of signs were perfected? (3) In those sciences in which there is general agreement as to the truth, is this because of the perfection of the signs employed in them? (4) In those branches of knowledge that provide inexhaustible fuel for dispute, is the division of opinion a necessary effect of the inexactitude of the signs employed in them? (5) Is there any means of correcting badly made signs and of rendering all sciences equally susceptible of demonstration? (*Mémoires de l'Institut National des Sciences et Arts. Sciences morales et politiques*, 1.i–ii)

Question (1) was on a thesis of *idéologie* itself, as may be seen in the passage from Cabanis quoted above. Degérando's answer was complex, but it was sufficiently affirmative to mark him an *idéologue*. On the one hand he felt that the mind needed no signs but merely an act of attention in order to pin down its sensations. On the other hand, "I shall give the name ['sign'] to every sensation that excites an idea in us in virtue of the association obtaining between them. Note carefully that it is not the sensation as such to which the name is given; it gets the name only in respect of the function it performs. Thus I shall say, for example, that the smell of a rose is the sign [not of the rose but] of the ideas of color and of form that the smell excites" (1.62–63). He distinguished between such prelinguistic signs and linguistic signs by pointing out that while the former "excite" ideas in us but attract attention to themselves, the latter lead our attention away from themselves to the ideas they have been made to signify, a formulation that constituted a refinement of the traditional distinction between natural and conventional signs.

In his detailed answers to questions (2) and (5) Degérando carefully criticized the many attempts at universal characteristics, calculi of reason, and philosophical languages that had been made by philosophers of the Renaissance and the Enlightenment. He laid down five criteria for such systems: (a) unambiguous relations between signs and signified ideas; (b) relations among signs exactly analogous to relations among signified ideas; (c) simplicity, that is, minimum number of primitives (*conditions premières*), each sign as abbreviated as possible, perspicuity of the sign system as a whole; (d) distinctness among signs of different sorts and among



syntactic relations of different sorts; (e) as many distinct sorts of signs as distinct sorts of ideas to be signified (4.353–355). The only hopes of satisfying such criteria, he maintained, lay along four different lines, or “systems for philosophical language.” Having examined each in detail, he concluded that all were in some respects unacceptable. Like Lambert, however, he suggested that a judicious application of the principles of such systems might produce some improvement in natural languages for philosophical purposes (4.355–415).

Perhaps even more important historically than his arguments against the feasibility of such artificial languages was his attack on the attitudes underlying them. It had usually been assumed that such a language would be international, and, indeed, if it were not, it would fail to achieve a good part of its purpose. But, Degérando maintained, there was no feasible means by which to establish it internationally, and even if it were established, it would soon be modified into separate dialects in various localities (3.557). Worst of all, the notion is pernicious, for such a language could at best be the instrument of communication exclusively among the learned and would thereby tend to separate them further from those they ought to instruct (3.572).

The fundamental mistake giving rise to all such schemes, according to Degérando, is the confusion of “the *method of reasoning* employed by the mathematicians with the *mechanical processes* of their calculations. Their method, as I have shown ..., they do have in common with the metaphysicians,” but the mechanical processes of their perfectly satisfactory artificial languages are the result of “the relative simplicity of the ideas on which they operate” (4.447–451). Other *idéologues*, particularly Destutt de Tracy, joined in this thoroughgoing repudiation of artificial languages for philosophy (see, for instance, *Mémoires de l’Institut*, Vol. III).

Questions (3) and (4) together called for an examination of Condillac’s contention that a science was to be identified with a well-made language. Cabanis, himself a scientist, and Destutt de Tracy had frequently made significant use of this doctrine, but Degérando rejected it in a way that seems symptomatic of the end of the Enlightenment conception of a science. Some of the basis for his answers to these questions is evident in his answers to questions (2) and (5). “A well-made language,” he maintained, “proclaims and presupposes a science that is already well advanced,” thus adopting the Cartesian position on this issue rather than the Leibnizian. “We shall say that the great art of perfecting a science consists above all in making better observations and only *then* adopting a

better language—i.e., one that is better suited to the observations that have been assimilated” (3.150–151). “The nomenclature of a science is related to the science itself as monuments are related to history: it preserves what is, but it can neither predict what is not yet nor unfold the future” (3.199).

Degérando resembled other *idéologues* more closely, however, in his view that improvements in philosophy—that is, in the analysis of sensations and ideas—did depend on a thorough examination of the natural language in which it was carried out. His own rather novel, never-realized scheme for accomplishing this was the construction of a philosophical dictionary.

It has been recognized that we can have clear ideas only in possessing a well-made language, and that a language can be well-made only in so far as we have reformed the most familiar operations of the mind from the very outset, only in so far as we have grasped the relation that interconnects them all. That being the case, we have felt the need of remaking the language in its entirety and, in some sense, recommencing the education of the human mind. The surest and perhaps the most truly efficacious means of accomplishing this great project would be, I think, the formation of a philosophical dictionary truly worthy of the name—one, that is to say, that would in some sense be a genealogical tree of our ideas and of the signs we use. Such a dictionary would be a sequence of definitions strictly bound to one another. Each notion would be defined in it by showing how it was acquired, or at least how it should have been acquired. The mind would find itself naturally led to *create* the words rather than seeking merely to *explain* them to itself.... The dictionary I propose would have an aim altogether different [from that of ordinary dictionaries]. In it one would seek to explain not so much how we speak as how we think; the conventions of the language would be presented in it as *results*, not as *principles*. ... [This dictionary would not be arranged in alphabetical order but] it would be a book, a history. The order of facts would be the only order observed in it. It would not ... be designed to be consulted occasionally, but it would have to be the object of a connected reading.... A definition would never be offered in it except in accordance with one general rule—that of determining an idea by means of tracing

it back to the ideas that must have preceded it in the age when language was instituted among men.... The dictionary would thus in some sense embrace the history of mankind and would serve as a natural introduction to all the sciences. The study of it would be necessary for all who wished to think well, and its formation would be one of the noblest undertakings of philosophy. (4.80ff.)

What little influence the *idéologues* had is somewhat more noticeable in British than in Continental philosophy of the nineteenth century, partly because of the interest of some of the Scottish commonsense philosophers in their work.

### UNIVERSAL GRAMMAR

Among the most important and distinctive influences on the philosophy of language of the Enlightenment was the development of universal grammar. In the broadest sense of the term it was, as defined by James Harris (1709–1780), “that grammar which without regarding the several idioms of particular languages only respects those principles that are essential to them all” (*Hermes, or a Philosophical Inquiry concerning Language and Universal Grammar*, 1751, Book I, Ch. 2). Although it resembles the speculative grammar of the Middle Ages in some of its basic assumptions, universal grammar seems to have had an independent origin that may with reasonable accuracy be dated 1660, the year in which Arnauld and Claude Lancelot published the *Port-Royal Grammar—Grammaire générale et raisonnée*.

Lancelot was a grammarian in the scholastic (rather than humanist) tradition who provided the subject matter that Arnauld presented in accordance with Descartes’s method, believing that he was thereby “developing in grammar a branch of Cartesianism” (p. 137). In grammar, as in every subject, the method consisted fundamentally in “beginning with the most general and simplest matters in order to proceed to the least general and most complex,” and in the study of language one therefore had to begin with principles and elements common to all languages in order to proceed to the study of one’s own and other particular languages. Thus, universal grammar, as Arnauld (and many of his successors) conceived of it, was an investigation of language (*langage*) designed as a propaedeutic to the study of languages (*langues*). In practice, however, the elements and principles of universal grammar tended to be those of traditional Latin grammar, and thus many of the so-called universal grammars,

at least before Condillac, are of little value either to linguists or to philosophers.

More important than the content of the early treatises on universal grammar, however, is the connection they established between grammar and philosophy, especially in France, where universal grammar dominated linguistic studies for 150 years following the *Port-Royal Grammar*. César Chesneau Dumarsais (1676–1756), the foremost of the universal grammarians between Arnauld and Condillac, maintained that “grammarians who are not philosophers are not even grammarians” (*Véritables Principes de la grammaire*, 1729, 1.201), and men engaged in the inquiry at that time styled themselves *grammairiens-philosophes*. Dumarsais seems often to have thought of “philosophy” in connection with grammar as no more than a certain scientific attitude (see, for instance, his article “Grammairien” in the *Encyclopédie*), but he also held some of the views that were to serve as the basis for a more strictly philosophical grammar, as can be seen in his explanation that “grammar has a necessary connection with the science of ideas and reasoning because grammar treats of words and their uses and words are nothing but the signs of our ideas and our judgments” (1.201).

The *grammairiens-philosophes* began to be more markedly philosophers than grammarians beginning with the articles on grammatical topics in the *Encyclopédie*. Although Dumarsais was in general charge of them, several were written by philosophers such as Voltaire, Diderot, and Turgot; and the articles as a group contain less information on the announced topics than they do discussions of philosophical questions more or less vaguely associated with those topics. Nicolas Beauzée (1717–1789), one of the authors of those articles, defended the new approach to grammar in his *Grammaire générale* (1767): “Why should one think metaphysics out of place in a book on universal grammar? Grammar ought to expose the foundations—the general resources and the common rules—of language, and language is the exposition of the analysis of thought by means of speech. No aim is more metaphysical or abstract than that” (*Préface*, p. xvii).

In the works of Condillac, the emphasis was no longer on the propriety of taking a philosophical approach to grammar but rather on the fundamental importance of universal grammar as an inquiry serving the purposes of philosophy itself. In the “Motif des études” introducing his course of studies for the prince of Parma, Condillac described grammar as “a system of words that represents the system of ideas in the mind

when we wish to communicate them *in the order and with the interrelations we apperceive.*” Consequently, as he remarked in the “Objet de cet ouvrage” preceding his *Grammaire* (1775), he regarded grammar “as the primary division of the art of thinking. In order to discern the principles of language we must observe how we think; we must seek those principles in the analysis of thought. But *the analysis of thought is quite complete in discourse*, with more or less precision depending on the greater or less perfection of languages and the greater or less exactness of mind on the part of those who speak them. This is what makes me think of languages as so many analytic methods.”

Condillac’s elevation of universal grammar to the status of a fundamental philosophical inquiry marked the beginning of a new phase in the development of universal grammar, but his view of it was no more than a natural consequence of the then well-established belief that the construction of an absolutely universal grammar for all languages was a feasible undertaking. If language depicted thought and all languages shared a set of elements and principles, then the study of those common elements and principles would provide a science of human thought.

For Condillac’s successors, the *idéologues*, who took the analysis of sensations and ideas to be the whole of philosophy, universal grammar became *the* philosophical method. As Destutt de Tracy put it, “this science may be called *idéologie* if one attends only to the subject-matter, *universal grammar* if one has reference only to the method, or *logic* if one considers only the goal” (*Éléments d’idéologie*, 1801, 1.5). It was in accordance with this conception of philosophy that the traditional chairs of logic and metaphysics in the *écoles centrales* of France were replaced in 1795 with chairs of universal grammar, which “by offering instruction in the philosophy of language would serve as an introduction to the course in private and public morality” (*Éléments, Préface*, p. xxiii).

Destutt de Tracy devoted the second part of his *Éléments d’idéologie*, some 450 pages, to a presentation of universal grammar suited to the purposes of the new course. Although he did occasionally cite parallel examples in other European languages and stress the value of knowing several languages, his principal interest was in what might fairly be described as the analysis of ordinary French. As an *idéologue* he was committed to provide analyses that would in every case disclose the signified idea (or sensation). His first step in establishing the conditions for such an analysis was to insist that the unit of signification was not the word or phrase, no matter how

complex, but only the proposition, the linguistic device expressive of a judgment. If one simply utters the words “Peter,” “to be not tall,” we say that it means nothing, it makes no sense, although if one merely changes the form of the verb so that one says “Peter is not tall,” thereby expressing a judgment, we can discern in what he says signs of his having an idea of Peter and an idea of his height (2.29–33).

Thus Destutt de Tracy committed himself to providing ideological analyses only if the linguistic entity to be analyzed occurs within a proposition. Even so, locating the signified idea sometimes required considerable ingenuity, as in his attempt to analyze all “conjunctions” in such a way as to show not only two propositions related by each conjunction but also the idea signified by each.

One can say as much regarding the conjunctions we use in asking questions, even though they might at first seem not to connect two propositions, because the first is suppressed. Thus when I say “how did you get in again?,” “why did you leave?,” I am really expressing these ideas: “I want to know [*Je demande*] *how* you got in again,” “I want to know *why* you left.” And when we unfold the sense of those conjunctions the result is: “I want to know *a thing that is the manner in which* you got in again,” “I want to know *a thing that is the reason for which* you left.” (2.136)

In Destutt de Tracy considerations of grammar were entirely subject to the demands of ideological analysis, as is plain not only in the structure of his final analyses above but also in his readiness, quite unusual even among *grammairiens-philosophes*, to revise the classifications of traditional grammar (treating the “adverbs” *comment* and *pourquoi* as “conjunctions”) when philosophical considerations seemed to call for their revision. When his volume devoted to universal grammar appeared in 1803, the experimental substitution of universal grammar for logic and metaphysics in the schools had already been abandoned, along with the *idéologues’* highest hopes for revolutionizing philosophy.

In Germany, Christian Wolff and Lambert had taken notice of universal grammar, but the movement had no appreciable impact on philosophy. In England it affected mainly the work of Harris, of James Beattie (1735–1803), and of John Home Tooke (1736–1812), all of whom developed universal grammar far less as a philosophical than as a philological inquiry. By the beginning of the nineteenth century, universal grammar was rapidly going out of fashion as a branch of philosophy, even in France, despite the last efforts of the *idéologues*.

## THE NINETEENTH AND TWENTIETH CENTURIES

### BENTHAM

Jeremy Bentham (1748–1832) did almost all his work in philosophy of language during the last twenty years of his life, primarily under the influence of Locke, the *idéologues*, and the universal grammarians. In a passage distinctly reminiscent of Locke's call for the development of "the doctrine of signs," Bentham expressed his own conviction that "a demand exists for an entirely new system of *Logic*, in which shall be comprehended a *theory of language*, considered in the most general point of view" (*Works*, edited by John Bowring, 8.119–120). His belief in the importance of a theory of language within a system of logic seems to have made an impression on J. S. Mill and may mark the beginning of the return to a view of the interrelations of logic and language more like that prevailing in the later Middle Ages than like that of the eighteenth century.

Universal grammar constituted a part of Bentham's plan for fulfilling the demand he had recognized, and his account of its subject matter is modeled explicitly on what he considered the "pioneering" work of Tooke (see, for instance, 8.187–188). Unlike Tooke, however, Bentham was inclined to consider it a branch of philosophical rather than philological inquiry and echoed the *idéologues* in his claim that within "the field of universal grammar it is not enough for a man to look into the books that are extant on the subject of grammar, whether particular or universal—he must look into his own mind" (10.193). He also followed the *idéologues*, Degérando in particular, in rejecting Condillac's view that languages and analytic methods were identifiable. On the one hand, he held, the analysis of experience on the most primitive level was dependent on the prelinguistic faculty of attention; on the other hand, "every name, which is not, in the grammatical sense, a *proper* name, is the sign and result" of an act of *synthesis* rather than of analysis (8.75; 8.121–126). Bentham did, however, cite Lavoisier's Condillac-inspired reform of the language of chemistry as a prime example of the practical value of the philosophy of language (3.273).

On more strictly semantic questions Bentham occasionally wrote as if he had simply absorbed and to some extent clarified the doctrines of Locke's Book III, but when his most distinctive refinements of Locke are brought together, they mark a genuine advance in the history of semantics. He was in general agreement with Locke that "language is the sign of thought, an instru-

ment for the communication of thought from ... the mind of him by whom the discourse is uttered [to another mind].... The immediate subject of a communication made by language is always the state of the speaker's mind." The crucial doctrine of immediate signification, which in Locke had been obscured by his vacillating treatment of it, was explicated by Bentham as follows:

In both these cases ["I am hungry," "That apple is ripe"], an object other than the state of my own mind is the subject of the discourse held by me, but in neither of them is it the *immediate* subject. In both of them the *immediate* subject is no other than the state of my own mind—an *opinion entertained by me in relation to the ulterior object or subject*. ... [Language] may be the sign of ... other objects in infinite variety, but of this object [the utterer's state of mind] it is always a sign, and it is only through this that it becomes the sign of any other object. (8.329–331)

Since, however, "communication may convey information purely, or information for the purpose of excitation" (8.301), the immediately signified state of the speaker's mind may be either "the state of the *passive* or *receptive* part of it, or the state of the *active* or *concupiscible* part" (8.329). Bentham described the use of language as a medium of communication as its "transitive" use. "By its *transitive* use, the collection of these signs is only the vehicle of thought; by its *intransitive* use, it is an instrument employed in the creation and fixation of thought itself." Consequently the transitive use of language "is indebted for its existence" to the intransitive use (8.228–229, 8.301).

Partly because he had begun with "thoughts" rather than Lockean ideas as the immediate significata of linguistic signs, and perhaps also because of the similar position taken in Destutt de Tracy's universal grammar, Bentham recognized not words but propositions as the elements of significance.

If nothing less than the import of an entire proposition be sufficient for the giving full expression to any [but] the most simple thought, it follows that, no word being anywhere more than a fragment of a proposition, *no word is of itself the complete sign of any thought*. It was in the form of entire propositions that when first uttered, discourse was uttered.... *Words may be considered as the result of a sort of analysis—a chemicological process for which, till of a*

comparatively much later period than that which gave birth to propositions, the powers of the mind were not ripe. (8.320–323; italics added)

“In language, therefore, the *integer* to be looked for is an entire proposition” (8.188).

Many of Bentham’s predecessors, but especially Locke, had inveighed against the philosophers’ tendency to “take words for things.” Bentham’s refinement and extension of this notion into a doctrine of “linguistic fictions” is his most distinctive contribution to philosophy of language. He began by taking the evidently unprecedented step of defining extralinguistic elements in terms of the functions of certain elements of language.

An *entity* is a denomination in the import of which every subject matter of discourse, for the designation of which the grammatical part of speech called a noun-substantive is employed, may be comprised.... A *real* entity is an entity to which, on the occasion and for the purpose of discourse, existence is really meant to be ascribed.... A *fictitious* entity is an entity to which, though by the grammatical form of the discourse employed in speaking of it, existence be ascribed, yet in truth and reality existence is not meant to be ascribed.

Thus the noun-substantive “motion” in “that body is in motion” is the name of a fictitious entity, since “this, taken in the literal sense, is as much as to say—Here is a larger body, called a motion; in this larger body, the other body, namely, the really existing body, is contained.” While he insisted that linguistic fictions stood in need of what he called exposition, he also maintained that they were contrivances “but for which language ... could not have existence” (8.195–199).

The mode of exposition to which linguistic fictions were to be subjected was called paraphrasis, which “consists in taking the word that requires to be expounded—viz the name of a *fictitious* entity—and, after making it up into a *phrase*, applying to it another phrase, which, being of the same import, shall have for its principal and characteristic word the name of the corresponding *real* entity” (8.126–127). Since all words designative of nonphysical entities involved linguistic fictions, most of the work of philosophy, Bentham thought, would consist in such exposition of language.

In his *Principles of Morals and Legislation*, Chapter X, Bentham recommended a method of starting philosophical inquiry that was later to be employed and advocated

by J. L. Austin. “I cannot pretend,” Bentham said of his catalogue of motives in that chapter, “to warrant it complete. To *make sure* of rendering it so, the *only* way would be to turn over the dictionary from beginning to end; an operation which, in a view to perfection, would be necessary for more purposes than this” (italics added).

## HUMBOLDT

The special historical importance of the work of Wilhelm von Humboldt (1767–1835) lies in the fact that it incorporates the transition from the eighteenth-century philosophy of language to the nineteenth-century science of linguistics. It does so not only in respect of the philosophical doctrines presented in it but also because Humboldt coupled those doctrines with empirical investigations of the sort he considered to be demanded by his philosophy of language.

His most important work—*Ueber die Kawi-Sprache auf der Insel Jawa* (published 1836–1839)—begins with the lengthy philosophical essay “Ueber die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts.” In it he developed his single most influential and original notion—that language is to be viewed not as a finished product but as a continuous process (*Sie selbst ist kein Werk, ergon, sondern eine Tätigkeit, energie*), as the totality of instances of speech (or of the understanding of speech). Written words constitute language only when they are read and to the extent to which they are understood (*Gesammelte Schriften*, edited by A. Leitzmann, 7[No. 1].46 ff.). The rules of syntax and the individual words of a language are, then, the products of analysis, having real existence only insofar as they are embodied in instances of actual speech. Thus, as Destutt de Tracy and Bentham had observed from other points of view, “we cannot possibly conceive of language as beginning with the designation of objects by words and thence proceeding to their organization. In reality, discourse is not composed from words that preceded it. On the contrary, the words issued from the totality of discourse” (7[No. 1].72 ff.; cf. 7[No. 1].143).

The essential role played by language in fixing and organizing thoughts had been recognized long before Humboldt, but he extended that recognition into the bold new doctrine that language activity was the medium of contact between the mind and reality. “Man lives with the world about him principally, indeed ... exclusively, as language presents it to him.” Humboldt felt that this conception of language held the solution to the post-Kantian problems regarding subjectivity and objectivity.

In speech the energy of the mind breaks a path through the lips, but its product returns through our own ears. The idea is [thus] translated into true objectivity without being withdrawn from subjectivity. Only language can do this.... [Moreover,] just as the particular sound mediates between the object and the man, so the whole language mediates between him and the nature that works upon him from within and without. He surrounds himself with a world of sounds in order to assimilate the world of objects. (7[No. 1]. 55ff.)

Somewhat as Hamann had done, Humboldt thus believed that philosophy reduced to the philosophy of language, and that he had “discovered the art of using language as a vehicle by which to explore the heights, the depths, and the diversity of the whole world” (letter to Wolfe, 1805).

The differences among natural languages were philosophically as well as scientifically important in Humboldt’s view, and he was opposed to the prevailing eighteenth-century type of universal grammar, which achieved its universality at the expense of linguistic differences that happened not to fit the grammatical schema adopted by the grammarian-philosopher. He proposed instead, and provided examples of, a genuinely comparative grammar, insisting that the comparative grammarian avoid adopting the grammar of Latin or of his native language as the schema within which to organize the forms of other languages (“Ueber das Entstehen der grammatischen Formen und ihren Einfluss auf die Ideenentwicklung,” in *Gesammelte Schriften* 4.285 ff.). Each natural language, he believed, was characterized by its own “inner form,” expressive of the psyche of the nation within which it had developed and which it bound together. The distinctive inner form manifested itself in the root words as well as in the patterns of word combinations peculiar to the language. This doctrine, which powerfully influenced the development of linguistics, was in Humboldt’s presentation of it little more than a consequence of the traditional semantic doctrine that speech reflected not objects but man’s view of objects, coupled with the novel romanticist conviction that the reactions of men to the world around them were not everywhere the same. Not only were the grammatical differences among languages to be respected and studied in their own right, but the separate vocabularies were also to be reexamined with a view to discovering not interlinguistic synonymy (which, strictly speaking, was illusory) but the nuances of mean-

ing that gave expression to different world views (7[No. 1].59 ff.; 89ff.; 190ff.).

Humboldt’s immediate influence was not on philosophers but on other founders of the science of linguistics, particularly on Franz Bopp (1791–1867). What influence his work eventually had on philosophy of language, at any rate outside Germany, seems to have been transmitted indirectly through the work of nineteenth- and twentieth-century linguists.

## JOHNSON

Alexander Bryan Johnson (1786–1867), the earliest American philosopher of language, was an isolated figure in the history of semantics. Locke and the Scottish common-sense philosophers strongly influenced his work, and he had learned something of the *idéologues* through Dugald Stewart’s account of them. He seems, however, to have had little or no knowledge of his other predecessors and contemporaries. Johnson’s work on language, published in three successive versions and under various titles in 1828, 1836, and 1854, went unnoticed for a hundred years and has had no appreciable influence since its republication during the 1940s. As the circumstances of his work would lead one to expect, it was unusual for its time both in its insights and in its mistakes.

The mistake that led to most of the others and to some of his principal insights as well occurred in his account of the semantics of words, in which he identified the signification(s) of a word with the thing(s) to which the word is applied. “Every word,” he argued, “is a sound, which had no signification before it was employed to name some phenomenon.” Consequently, “words have no inherent signification, but as many meanings as they possess applications to different phenomena. *The phenomenon to which a word refers, constitutes in every case, the signification of the word*” (*Treatise on Language*, Lectures VI and V; italics added).

The phenomena available as referents (or meanings) were exhaustively divided by Johnson into “sights, sounds, tastes, feels, smells, internal feelings, thoughts, and words” (Lecture XI). The word *table*, for example, signifies both a sight and a feel, “two distinct existences” bearing a single name. In this way “language implies a oneness to which nature conforms not in all cases,” and men are prone to “make language the expositor of nature, instead of making nature the expositor of language” (Lecture III). Johnson made this common human failing his constant theme and provided several examples of philosophical and scientific difficulties that he felt were obviated by exposing a confusion of this sort as their source.

Philosophers, he suggested, might append to every “nominal unit that aggregates objects generically different” a capital letter—for instance, *S*, sight; *F*, feel—indicative of the phenomenon signified on each occasion. By that means David Hume, for example, might be seen to be announcing an “unconscious quibble,” when he says, “The table (*S*) which we see, seems to diminish (*S*) as we recede from it, but the real table (*F*) suffers no diminution (*F*).” The whole zest of the proposition consists in the sensible duality of each of the nominal units table and diminution.... We play bo-peep with words, by neglecting to discriminate the intellectually conceived oneness of diminution, and its physical duality” (*The Meaning of Words*, pp. 89–92).

In his account of the semantics of propositions Johnson remained faithful to the identification of meaning and referent with disastrous results, the most obvious of which was the confusion of meaningfulness (or meaninglessness) with truth (or falsity). “No proposition,” he held, “can signify more than the particulars to which it refers” (Lecture VIII). He saw that one difficulty with this doctrine was that under it “the proposition that all men must die seems equivalent only to the proposition that all men have died.” In his attempt to preserve the “universal application” of such general propositions, he adopted the indirect criterion of the failure of their negations to refer to any sensible particular. Thus “the proposition that all men will die, possesses a universal application for the reason that to say, some men will not die, *refers to no sensible particulars, and hence is insignificant*” (Lecture IX; italics added).

It was, however, this same approach to the semantics of propositions that led Johnson to develop and make critical use of a verifiability criterion of meaningfulness. Chemists, he remarked, had an indisputable right to “say simply that they can produce hydrogen gas, and oxygen, from water, and vice versa,” but what they say instead is “that water is nothing but a combination of these gases. The assertion is true, *so long as it means [merely] the phenomena to which it refers*; but it produces wonder, because we suppose it has a meaning beyond the phenomena” (Lecture VII; italics added). Similarly, “if you inquire of an astronomer whether the earth is a sphere, he will desire you to notice what he terms the earth’s shadow in an eclipse of the moon, the gradual disappearance of a ship as it recedes from the shore, &c. After hearing all that he can adduce in proof of the earth’s sphericity, *consider the proposition [‘the earth is a sphere’] significant of these proofs. If you deem it significant beyond them, you are deceived by the forms of language*” (Lecture VIII; italics

added). In his verifiability criterion of meaningfulness and in his related discrimination of significant and insignificant questions (Lectures XIX ff.), Johnson anticipated some of the fundamental semantic principles of the pragmatists and positivists.

## MILL

Many of the remarkable developments in semantics in the late nineteenth and early twentieth centuries took place under the influence of or in reaction against the doctrines of John Stuart Mill (1806–1873). He presented his “philosophy of language” (a designation he seems to have made current) in Book I and Chapters III–VI of Book IV of *A System of Logic* (1843), acknowledging the influence of the medieval logicians and Hobbes in particular, but also of Locke, Dugald Stewart, and others in the tradition of British empiricism. Like most of his empiricist predecessors in France and England, Mill believed that a philosophical inquiry into language had a high therapeutic value for philosophy itself, viewing metaphysics as “that fertile field of delusion propagated by language” (1.7.5).

By way of explaining his return to the practice of associating semantical inquiries with logic, Mill argued that since “language is an instrument of thought,” not only in the reasoning process proper but in the antecedent operations of classification and definition, “logic ... includes, therefore, the operation of Naming” (introduction, Sec. 7). It is not clear whether Mill intended to *identify* the ratiocinative use of language with naming or to claim that all language *stems from* the operation of naming, but he did revert to the tradition of considering “names” as the elements of his semantic theory. And since he took it to be obvious that “a proposition ... is formed by putting together two names” (1.1.2), it seemed equally obvious that “the import of words [or names] should be the earliest subject of the logician’s consideration: because without it he cannot examine into the import of propositions” (1.1.1).

NAMES. In his account of the import of names, Mill began by taking the unusual tack of defending “the common usage” against the view of “some metaphysicians,” arguing that words are “names of things themselves, and not merely of our ideas of things.” (Although there are passages in Hobbes and Locke, for example, that can be interpreted as expressions of that view, neither they nor, it seems likely, anyone else held quite the view Mill was criticizing.) “It seems proper,” Mill claimed, “to consider a word as the *name* of that ... concerning which, when we

employ the word, we intend to give information.” When, however, “I use a name for the purpose of expressing a belief, it is a belief concerning the thing itself, not concerning my idea of it,” even when the belief in question is one concerning some idea of mine (1.2.1).

A name, in Mill’s adaptation of scholastic terminology, was said to *denote*, individually and collectively, the things of which it was the name, “the things of which it can be predicated.” But, as Mill observed, “by learning what things it is a name of, we do not learn the *meaning* of the name” (1.2.5). A name happens “to fit” a given thing “because of a certain *fact*. ... If we want to know what the fact is, we shall find the clue to it in the *connotation*” of the name (1.5.2). The connotation of the name is the “attribute” or set of attributes possession of which by a given thing is the fact in virtue of which the name fits the thing, and “the *meaning* of all names, except proper names [which have no meaning] and that portion of the class of abstract names which are not connotative [such as ‘squareness,’ which *denotes* a single attribute], resides in the connotation” (1.5.2). Mill recognized the connection of this distinction with the doctrine of denominatives (see the discussion of Anselm above), and in a note to 1.5.4 he indicated its relations to Hamilton’s intension-extension distinction (see below).

**CONNOTATION AND DENOTATION.** Mill believed that the connotation-denotation distinction was “one of those which go deepest into the nature of language” (1.2.5). He made considerable use of it himself, and it played an important part in philosophical discussions for at least seventy-five years afterward. It is, however, a notoriously unclear distinction, especially in Mill’s own treatment of it. With regard to denotation, for example, he claimed that a “concrete general name” such as “man” denotes Socrates—that is, is a name of, is predicable of that individual—but he claimed also that it denotes *the class of which that individual is a member, which (at best) introduces a crucial ambiguity into the notion of denotation.* With regard to connotation, the most serious difficulty centers on the notion of “attributes,” which Mill suggested at one point was to be identified with what medieval logicians meant by “forms” (1.2.5n). In an evidently more careful account he declared that “the meaning of any general name is some outward or inward phenomenon, consisting, in the last resort, of feelings; and these feelings, if their continuity is for an instant broken, are no longer the same feelings, in the sense of individual identity.

What, then, is the common something which gives a meaning to the general name? Mr. [Herbert] Spencer can only say, it is the similarity of the feelings: and I rejoin, the attribute is precisely that similarity.... The general term *man* does not connote the sensations derived once from one man.... It connotes the general type of the sensations derived from all men, and the power ... of producing sensations of that type” (2.2.4n). The only plausible interpretation of this doctrine seems to bring it very close to Hobbes’s (or Locke’s) actual account of words as signs of our ideas (despite Mill’s attack on its weakest version: Words are *names* of our ideas), for in the end Mill’s semantics of words appears to be founded on the familiar view that words are signs of extramental entities (denotation) only in virtue of being signs of mental entities of some sort (connotation). Mill surely would have recoiled at the suggestion that his doctrine of the connoted attribute as a “general type” of sensations committed him to an acceptance of extramental metaphysical entities.

After a detailed, ingenious investigation of the semantics of many-worded connotative concrete individual names (1.2.5), frequently discussed by his successors, Mill turned to the semantics of propositions. His account of the meaning of names and his view that the meaning of a proposition is a function of the meanings of the names that serve as its terms led naturally to his view that “when ... we are analyzing the meaning of any proposition in which the predicate and the subject, or either of them, are connotative names, it is to the *connotation* of those terms that we must exclusively look, and not to what they *denote*.” The view of Hobbes—that the predicate term is to be considered a name of whatever the subject term names—“is a mere consequence of the conjunction between the two attributes,” the connotations of the two terms, and is adequate only in case both terms are nonconnotative names (1.5.2).

Thus, “all men are mortal” asserts that “the latter set of attributes *constantly accompany* the former set.” And on the basis of the account of attributes introduced above, “we may add one more step to complete the analysis. The proposition which asserts that one attribute always accompanies another attribute, really asserts thereby no other thing than this, that one phenomenon always accompanies another phenomenon; in so much that where we find the latter, we have assurance of the existence of the former.” He was, however, careful to note that “the connotation of the word *mortal* goes no farther than to the occurrence of the phenomenon at some time or other” (1.5.4).



When he came to discuss “real” (as opposed to “verbal”) propositions, however, Mill disclosed that with respect to real propositions the account just cited was only one of “two formulas” in which “their import may be conveniently expressed.” The account in terms of companion sets of attributes is suited to the view of real propositions “as portions of speculative truth.” But they may be viewed also “as memoranda for practical use,” and Mill’s consideration of them in this light prefigured some elements of pragmatist theories of meaning. “The practical use of a proposition is, to apprise or remind us what we have to expect in any individual case which comes in the assertion contained in the proposition. In reference to this purpose, the proposition, All men are mortal, means that the attributes of man are *evidence of*, are a *mark of*, mortality; ... that where the former are we ... [should] expect to find the latter.” The two formulas for expressing the import of real propositions are, Mill maintained, “at bottom equivalent; but the one points the attention more directly to what a proposition means, the latter to the manner in which it is to be used” (1.6.5).

Mill agreed with the majority of his philosophical contemporaries in deploring attempts to devise a formalized language for philosophy and suggesting that philosophers reform the natural languages for their uses. He was in a minority, however, in urging philosophers to have a healthy respect for natural languages. One of the “inherent and most valuable properties” of a natural language is “that of being the conservator of ancient experience”—“Language is the depository of the accumulated body of experience to which all former ages have contributed their part, and which is the inheritance of all yet to come.” Consequently, “it may be good to alter the meaning of a word, but it is bad to let any part of the meaning drop” (4.4.6). Mill was emphatic about the special respect with which words of uncertain connotation were to be treated, and he laid down as a principle for the guidance of philosophers that “the meaning of a term actually in use is not an arbitrary quantity to be fixed, but an unknown quantity to be sought” (4.4.3). The attitude toward natural languages enjoined on philosophers by Mill was in part the attitude adopted by J. L. Austin and other twentieth-century philosophers of ordinary language.

## PEIRCE AND THE PRAGMATISTS

In a tradition stemming from Locke, Charles Sanders Peirce (1839–1914) characterized logic “in its general sense” as “*semiotic* (σημειωτική), the quasi-necessary, or formal, doctrine of signs” (*Collected Papers* 2.227) and went much further than anyone before him had tried to

go toward the development of a completely general theory of signs. (Insofar as Peirce’s semiotic deals with non-linguistic signs, it lies outside the scope of this article, but his elaborate, varying terminology makes it difficult to present a single standard version of even that portion of the theory which is directly relevant to his treatment of linguistic meaning.)

Peirce seems sometimes to have thought of semiotic as a generalized version of the medieval trivium, describing its three branches as “pure grammar,” “logic proper,” and “pure rhetoric” (2.228–229). The first branch was to investigate the necessary conditions of meaningfulness, the second was to investigate the necessary conditions of truth, and the third was “to ascertain the laws by which in every scientific intelligence one sign gives birth to another, and especially one thought brings forth another” (2.229). These branches, with their subject matter somewhat differently described, were to become well known in twentieth-century philosophy under the designations “syntactics,” “semantics,” and “pragmatics” respectively—designations introduced by Charles W. Morris (*Foundations of the Theory of Signs*, 1938) and used extensively by Rudolf Carnap and others.

“Semiosis” was Peirce’s name for an instance of signification, which he described as involving three principal elements: the *sign*, “something which stands to somebody for something in some respect or capacity” (its “ground”); the *object*, that for which the sign stands; and the *interpretant*, another sign, equivalent to or “more developed” than the original sign and caused by the original sign in the mind of its interpreter (2.228). The notion of the interpretant is the distinctive element in Peirce’s general account of signification and the one that played the central role in his pragmatism (or “pragmaticism”), which he often described as consisting entirely in “a method for ascertaining the real meaning of any concept, doctrine, proposition, word, or other sign” (5.6).

Some of Peirce’s predecessors had already suggested that the meaning of a word could be determined only on a given occasion of its occurrence within a propositional context, but in Peirce’s the traditional primacy of the semantics of words over the semantics of propositions was so thoroughly overturned that his theory of linguistic meaning is almost exclusively a theory regarding the meaning of whole propositions. According to that theory, a proposition, like every other sign, has an object—some state of affairs, factual or otherwise. The *meaning* of a proposition, however, he identified not with its object but with one particular kind of effect of the proposition on

an interpreter, namely, its “*logical*” (as opposed to “emotional” or “energetic”) *interpretant* (5.476).

Peirce’s definitive account of the logical interpretant appeared in the 1905 paper “What Pragmatism Is” (5.411–434), in which he attempted as well to explain the distinctive and often misinterpreted “futuristic” aspect of pragmatist meaning theory.

The rational meaning of every proposition lies in the future. How so? The meaning of a proposition [that is, its logical interpretant] is itself a proposition. Indeed, it is no other than the very proposition of which it is the meaning; it is a translation of it. But of the myriads of forms into which a proposition may be translated, what is that one which is to be called its very meaning? It is, according to the pragmatist, that form in which the proposition becomes applicable to human conduct, ... that form which is most directly applicable to self-control under every situation and to every purpose. This is why he locates the meaning in future time; for future conduct is the only conduct that is subject to self-control.

The only form of the proposition that would satisfy all these conditions was “the general description of all the experimental phenomena which the assertion of the proposition virtually predicts. For an experimental phenomenon is the fact asserted by the proposition that action of a certain description will have a certain kind of experimental result; and experimental results are the only results that can affect human conduct.” Thus, as Peirce finally conceived of it, the meaning of a proposition is evidently to be explicated in the form of a true conditional with the original proposition as antecedent and, as its consequent, a conjunction of propositions constituting “the general description of all the experimental phenomena which the assertion of the [original] proposition virtually predicts.”

Among the more striking problems in this account are (1) the difficulty of applying it to propositions other than those which occur within the context of an experimental science and (2) the fact that the meaning of a proposition is said to consist in other propositions, the meanings of which are presumably explicable in the same fashion, *ad infinitum*. Peirce was aware of both these problems. His response to (1) was generally to minimize the differences between the context of an experimental science and other contexts within which propositions occur, although he did occasionally, especially in his later writings, acknowledge the perhaps insuperable difficul-

ties in employing this as a completely general theory of linguistic meaning. With regard to (2) Peirce was at first inclined to claim that a proposition (or any other sign) was, indeed, imperfectly significant if the series of its interpretants was finite (“Sign,” in Baldwin’s *Dictionary*).

Later, however, the notion of “the *ultimate* logical interpretant” was introduced. “The real and living logical conclusion” of the series of logical interpretants is an expectation (on the interpreter’s part) of certain phenomena “virtually” predicted by the assertion of the original proposition. This expectation Peirce frequently referred to as “habit.” “The deliberately formed, self-analyzing habit—self-analyzing because formed by aid of analysis of the exercises that nourished it—is the living definition, the veritable and final logical interpretant” (5.491; cf. 5.486). Habit, which Peirce sometimes described as a “readiness to act in a certain way under certain circumstances and when actuated by a given motive,” was not itself a sign and so stood in no need of interpretants of its own.

It was on this very point that Peirce thought his own doctrine differed from that of William James (1842–1910). “In the first place,” he wrote, “there is the pragmatism of James, whose definition differs from mine only in that he does not restrict the ‘meaning,’ that is, the ultimate logical interpretant, as I do, to a habit, but allows percepts, that is, complex feelings endowed with compulsiveness to be such” (5.494). James’s own definition of “pragmatism” in Baldwin’s *Dictionary* identified it as “the doctrine that *the whole meaning* of a conception *expresses itself in practical consequences* either in the shape of conduct to be recommended or in that of experiences to be expected, if the conception be true” (italics added), but in doing so he evidently believed he was promulgating “Peirce’s principle ... that *the effective meaning* of any philosophic proposition can always be *brought down to some particular consequence*, in our *future practical experience*, whether active or passive” (*Collected Essays and Reviews*, edited by R. B. Perry, p. 412; italics added). James’s conception of pragmatism as a theory of meaning (and of truth) was, however, unquestionably broader and less carefully qualified than Peirce’s and may fairly accurately be summarized in his own characteristic observation that concepts and propositions “have, indeed, no meaning and no reality if they have no use. But if they have any use they have that amount of meaning. And the meaning will be true if the use squares well with life’s other uses” (*Pragmatism*, p. 273).

Pragmatism first became generally known in the form given it by James and in the still wider “humanism”

of F. C. S. Schiller, and it was in those forms that it was subjected to intense criticism at the beginning of the twentieth century by F. H. Bradley and G. E. Moore, among others. Peirce's more intricate and interesting theory of meaning was not really considered in its own right until some years afterward, perhaps beginning with the publication of C. K. Ogden and I. A. Richards's very influential *The Meaning of Meaning* in 1923, in which some ten pages were devoted to an exposition of Peirce's semiotic.

At the same time a pragmatist theory of meaning more complex and no less broad than James's was being developed in the "instrumentalism" of John Dewey (1859–1952). Dewey discussed meaning of every imaginable sort and in countless contexts, with the result that it is difficult to elicit from his many writings a genuinely representative doctrine specifically of linguistic meaning. Perhaps the least misleading single source is Chapter 5 of his *Experience and Nature*, first published in 1925. His position there was as follows:

The sound, gesture, or written mark which is involved in language is a particular existence. But as such it is not a *word*, and it does not become a word by declaring a mental existence; it becomes a word by gaining meaning; and it gains meaning when its use establishes a genuine community of action.... Language and its consequences are characters taken on by natural interaction and natural conjunction in specified conditions of organization.... Language is specifically a mode of interaction of at least two beings, a speaker and a hearer; it presupposes an organized group to which these creatures belong, and from whom they have acquired their habits of speech. It is therefore a relationship, not a particularity.... The meaning of signs moreover always includes something common as between persons and an object. When we attribute meaning to the speaker as *his* intent, we take for granted another person who is to share in the execution of the intent, and also something, independent of the persons concerned, through which the intent is to be realized. Persons and thing must alike serve as means in a common, shared consequence. This community of partaking is meaning.

Even when, as in these passages, Dewey seems to have been considering linguistic meaning specifically, there is a real possibility that his intentions were much broader, for his conception of language was itself considerably

broader than that of most philosophers. He was, for example, prepared to say that "because objects of art are expressive, they are a language. Rather they are many languages. For each art has its own medium and that medium is especially fitted for one kind of communication.... The needs of daily life have given superior practical importance to one mode of communication, that of speech" (*Art as Experience*, 1935, p. 106).

Pragmatist theories of meaning, beginning with Peirce's 1878 paper "How to Make Our Ideas Clear," are alike in little more than their tendency to associate the meaning of a proposition with the conditions of its verification, but in that respect they may be said to have inaugurated twentieth-century developments of empiricist and operationalist theories of meaning.

## FREGE

The contributions of Gottlob Frege (1848–1925) to logic, philosophy of mathematics, and semantics were largely unappreciated at the time of their publication, primarily during the last quarter of the nineteenth century. Their influence (direct or indirect) on recent philosophy has been so great, however, that Frege might fairly be characterized as the first twentieth-century philosopher. In his *Begriffsschrift* (1879) he developed "a formalized language of pure thought modeled on the language of arithmetic," which has been recognized as the first really comprehensive system of formal logic. In his other two major works, *Die Grundlagen der Arithmetik* (1884) and *Die Grundgesetze der Arithmetik* (1893–1903), he tried to show "that arithmetic is founded solely upon logic."

Philosophical problems encountered by Frege in those highly technical undertakings were explored by him in several papers that have had a wider influence than his books have had. As the topics of the books might lead one to expect, his philosophical papers are concerned almost exclusively with one or another aspect of systems of signs. Of these papers, the one that has had most effect on the development of semantics is "Ueber Sinn und Bedeutung" (1892, translated in P. T. Geach and Max Black, *Translations from the Philosophical Writings of Gottlob Frege*, 1952), although the doctrine presented in it may prove to be historically less important than the doctrine of "functions" developed in other papers.

There is a broad and not wholly misleading similarity between Frege's distinction of sense (*Sinn*) and reference (*Bedeutung*) and such distinctions as comprehension-extension (Arnauld), intension-extension (Hamilton), connotation-denotation (Mill), depth-breadth (Peirce). It seems possible, however, that Frege

developed his distinction independently; in any case, the details of his doctrine are quite novel. Most important, perhaps, was his discovery of special contexts rendering the application of any such distinction problematic.

**SENSE AND REFERENCE.** Frege's development of the doctrine of sense and reference began, characteristically, in a consideration of the relation of identity: "=". He noted that " $a = a$  and  $a = b$  are obviously statements of different cognitive value," which they would not be if we were to take the relation to hold "between that which the names '*a*' and '*b*' designate" or refer to (Geach and Black, p. 56). Consequently, "it is natural now, to think of there being connected with a sign (name, combination of words, letter), besides that to which the sign refers, which may be called the reference of the sign, also what I should like to call the *sense* of the sign, wherein the mode of presentation is contained.... The reference of 'evening star' would be the same as that of 'morning star,' but not the sense" (p. 57).

Frege first applied his distinction to proper names, by which he meant any "designation of a single object." In keeping with Arnauld's similar distinction but in opposition to Mill's, Frege ascribed sense as well as reference to such designations. "A proper name (word, sign, sign combination, expression) *expresses* its sense, *stands for* or *designates* its reference. By means of a sign we express its sense and designate its reference" (p. 61). The sense of "Aristotle" might, for instance, be taken to be the following: the pupil of Plato and the teacher of Alexander the Great" (p. 58n).

Certain expressions, such as "the least rapidly convergent series," have a sense, he maintained, but no reference at all. "In grasping a sense, one is not certainly assured of a reference" (p. 58). An expression that has a reference "must not be taken as having its ordinary reference" when "standing between quotation marks" (pp. 58–59). Such observations had been made before, but Frege seems to have been the first to try to show what that extraordinary reference might be and, more important, to recognize that many different linguistic contexts affected the reference of expressions included within them, especially indirect discourse and subordinate clauses following such verbs as "hear," "conclude," "perceive," and "know." He claimed, for example, that "in reported speech, words ... have [not their customary but] their *indirect* reference," and that "the indirect reference of a word is ... its customary sense" (p. 59). His account of the effect of such contexts on reference has not been widely accepted, but the problems raised by it have stimulated

the widespread interest of twentieth-century philosophers in such now familiar topics as synonymy, opacity of reference, Leibniz's Law, and what, following Franz Brentano (see below), have come to be called intentional contexts.

Frege was concerned with saying what sort of entities sense and reference were. In the case of a proper name his description of the reference was relatively unproblematic: "a definite object (this word ['object'] taken in the widest range)"—(p. 57)—so wide that " $2 + 2$ " and "4," for example, were two proper names with one and the same "object" as their reference. Regarding the sense of a proper name, he found it easier to say what it was not: "The reference of a proper name is the object itself which we designate by its means; the idea, which we have in that case, is wholly subjective; in between lies the sense, which is indeed no longer subjective like the idea, but is yet not the object itself" (p. 60). Thus, there is "an essential [subjective-objective] distinction between the idea and the sign's sense." Frege seems not to have completely depsychologized the notion of the sense of the sign, however, since he suggested that it may be an element in humankind's "common store of thoughts which is transmitted from one generation to another" rather than "a part or a mode of the individual mind" (p. 59).

In Frege's discussion of the sense and reference of declarative sentences, the doctrine of the sense was relatively straightforward while the account of the reference became problematic. A sentence, he held, "contains a thought," and by "a thought" he meant "not the subjective performance of thinking but its objective content, which is capable of being the common property of several thinkers" (p. 62 and note). The two sentences "the morning star is a planet" and "the evening star is a planet" contain different thoughts, as may be seen from the fact that "anybody who did not know that the evening star is the morning star might hold the one thought to be true, the other false. The thought, accordingly, cannot be the reference of the sentence, but must rather be considered its sense" (p. 62). We are content to consider only the sense of sentences as long as we are not concerned to judge of their truth or falsity, but "in every judgment, no matter how trivial, the step from the level of thoughts to the level of reference (the objective) has already been taken" (p. 64). What we seek in judgment is the truth-value of the sentence. "We are therefore driven into accepting the truth-value of a sentence as constituting its reference.... Every declarative sentence concerned with the reference of its words is therefore to be regarded as a proper name,

and its reference, if it has one, is either the True or the False” (p. 63).

As his use of the phrase “driven into accepting” indicates, Frege was well aware that this was a startling doctrine of the semantics of sentences. Much of the remainder of his paper on sense and reference was devoted to considerations that he felt tended to support it, among them Leibniz’s Law; for “what else but the truth-value could be found, that belongs quite generally to every sentence if the reference of its components is relevant, and remains unchanged by substitutions of the kind in question?” (p. 64).

The doctrine that sentences are proper names, whether or not of the True and the False, had an important negative effect in that its rejection by Wittgenstein (see, for instance, *Tractatus* 3.143) and by Bertrand Russell under Wittgenstein’s influence (see, for instance, “Logical Constructions,” Lecture I) helped to shape the course of philosophy of language in the twentieth century.

It is quite likely, however, that Frege’s assimilation of declarative sentences to proper names was not quite so thorough or simple as his presentation of it in “Ueber Sinn und Bedeutung” suggests. Some of his remarks in an earlier paper, “Funktion und Begriff” (1891), at least raise the possibility that he may have denied proper-name status to sentences actually being used in making assertions (rather than considered as examples). In order to make what he took to be the indispensable “separation of the act from the subject-matter of judging” he introduced his assertion sign—“┊”—“so that, e.g., by writing

$$\vdash 2 + 3 = 5$$

we assert that  $2 + 3$  equals 5. Thus here we are *not just writing down a truth-value*, as in

$$2 + 3 = 5,$$

but also at the same time *saying that it is the True*.” And in a note to this passage he maintained that “‘┊  $2 + 3 = 5$ ’ does not *designate* [that is, refer to] anything; it *asserts something*” (p. 34; italics added).

## MAUTHNER

Of the several late nineteenth-century philosophers writing in German whose work centered on a concern with language, the most unusual was Fritz Mauthner (1849–1923). His principal work, *Beiträge zu einer Kritik der Sprache*, fills three large volumes and went through three editions, the first in 1901–1902. In his thoroughgoing

attempt to transform all philosophy into philosophy of language, in his criticisms of Kant, and in his penchant for paradox he resembled Hamann, whom he admired, and also, to some extent, Humboldt. He seems, however, to have been most powerfully influenced by the positivism of Ernst Mach and especially by Hume’s skepticism, adopting as his philosophical watchword “Back to Hume!”

Part of what Mauthner meant by that is apparent in the epistemological doctrine on which he founded his critique of language: “Our memory [with which he identified our knowledge] contains nothing but what our poor fortuitous senses [*Zufallsinne*] have presented to it” (*Beiträge* 3.536). By calling our senses “fortuitous” he was calling attention to the fact that if we had been otherwise equipped with senses, we might have framed a very different view of the world. Language, however, depicts not the world but a world view. Therefore, any attempt to infer propositions regarding reality from facts of language is a form of “word-superstition.”

Moreover, each man’s individual senses present a world view unique in certain ultimately undeterminable respects, and so communication by means of language, even if it purports to be no more than an exchange of views, is fundamentally illusory. “No man knows the others.... With respect to the simplest concepts we do not know of one another whether we have the same representation associated with one and the same word.”

From such avowedly Lockean observations Mauthner drew the typically paradoxical conclusion that “by means of language men have made it forever impossible to get to know one another” (1.54). Thus he characterized language as “nothing other than just the community or the mutuality of world-views.” It is not a tool for the communication or acquisition of knowledge; indeed, it is not a tool or an object of any sort but merely a practice, a use. And “because it is no object of use but use itself, it perishes without use” (1.24). But of all Mauthner’s many characterizations of language the one most suggestive of distinctively twentieth-century attitudes is this: “Language is merely an apparent value [*Scheinwert*], like a rule of a game [*Spielregel*], which becomes more binding as more players submit to it, but which neither alters nor comes into contact with [*begreifen*] the world of reality” (1.25).

Philosophy, in Mauthner’s view, had to become a critique of language if it was to be anything at all, and in that guise its principal function was to be therapeutic. “Philosophy ... cannot wish to be anything more than critical attention to language. Philosophy can do no more with

respect to the organism of language or of the human spirit than can a physician with respect to the physiological organism. It can attentively observe and designate the developments with names" (1.657). "If I want to ascend into the critique of language, which is at present the most important undertaking of thinking mankind, I must do away with language behind me and before me and in me, step by step—I must break in pieces each rung of the ladder as I tread on it" (1.1–2). It comes as no surprise to learn that the end of this therapeutic process was to be silence, the silence of mystical contemplation.

In the course of the long process, however, Mauthner found occasion to make many insightful observations on traditional problems of the philosophy of language. As several of his contemporaries and immediate predecessors had done, he recognized not the word but the sentence (*Satz*) as the unit of meaningfulness and described the meaning of the word as a function of its use in a given sentence. Another position that was not new but to which he gave an especially forceful presentation was the rejection of the view "that because there is a word, it must be a word *for* something; that because a word exists there must exist something real corresponding to that word." This form of word superstition he regarded as "mental weakness" (2nd ed., 1.159).

It is probably only coincidence, but the name theory of linguistic meaning against which Mauthner inveighed bears a strong resemblance in some respects to the theories of Alexius Meinong and Edmund Husserl then being published and to the early views of Russell and Wittgenstein. Mauthner also opposed efforts at universal grammar (such as some of Brentano's followers were then engaged in) and mathematical logic, maintaining that all formalization of language obliterated or obscured far more than it clarified. Thus, he noted that "if someone says 'cheese is cheese' ... this utterance is *not* an instance of the general formula 'A = A'" (3.366), a formula "so empty that outside logic it must arouse the suspicion of insanity" (*Wörterbuch der Philosophie*, article "A = A").

Perhaps more than any other philosopher of language Mauthner had an appreciation of the history of the subject; at one time, in fact, he planned a fourth volume of his *Beiträge* that was to present the approach to the critique of language throughout the history of philosophy. Even as it stands, however, his work is filled with references to his predecessors and evaluations of their work from the viewpoint of the critique of language. Aristotle, for example, comes off badly, but Locke ranks very high. Indeed, Mauthner took "the English" to task for abandoning the work of Locke, for failing to see that "the con-

tent of their famous 'understanding' is simply the dictionary and grammar of human language" (*Beiträge* 3.535).

Mauthner's own effect on the history of the philosophy of language is still difficult to assess. Wittgenstein certainly knew of his work (see, for instance, *Tractatus* 4.0031). Whether or not Wittgenstein's turn in the direction of some of Mauthner's doctrines in *Philosophical Investigations* was coincidence or derived in part from Mauthner's influence remains an open question.

## HUSSERL AND MEINONG

The students of Franz Brentano (1838–1917), among whom were Husserl, Meinong, Anton Marty, and Kazimierz Twardowski, were alike at least in taking Brentano's concept of intentionality as a point of departure in their own philosophizing. Brentano had introduced intentionality in his *Psychologie vom empirischen Standpunkt* (1874) as the differentia of "mental states," a characteristic "which the schoolmen of the middle ages called the intentional (or mental) inexistence of an object and which we ... describe as *the relation to a content*, or *the direction to an object* (by which we need not understand a reality), or an *immanent objectivity*. Every mental state possesses in itself something that serves as an object, although not all possess their objects in the same way" (*Psychologie* 2.1.5; italics added).

The "intending" of an object by a mental state, the "directedness" of a mental state, bears a close enough resemblance to what is called significance in other contexts that much of what Brentano and his followers had to say in working out their central doctrine of intentionality has some relevance to semantics, broadly conceived. More specifically, the notion of intentionality underlies the considerable discussion in semantics of "intentional contexts," produced as a result of the ordinary use of such "intentional words" as "believe," "want," and "ascribe." For present purposes, however, our attention is confined to what Brentano's two best-known students, Husserl and Meinong, had to say expressly about language and linguistic meaning. The doctrines of both men passed through several stages of development and contain many complexities, only a few of which can be noted here.

The philosophy of language of Edmund Husserl (1859–1938) was developed at various places in his work but is concentrated in the first and fourth essays in his *Logische Untersuchungen* (1900–1901; rev. ed., 1913–1921). The first, titled "Expression and Meaning" (2.23–105), was designed partly as a general preparation for intensive work in phenomenology as conceived by

Husserl. It opens with an investigation of signs in general and proceeds to the consideration of expressions, signs that may be said to have meanings (*Bedeutungen*) and not merely to indicate something. The three ingredients of meaningfulness, or “the meaning-situation” are (1) a “meaning-endowing act,” or “meaning-intention” on the part of the producer of the expression, which may be associated with a “meaning-fulfilling act” on the part of an interpreter of the expression; (2) the *content* of these acts, or the *meaning* of the expression; (3) the *object* of these acts, or, in Husserl’s broader terminology, the *objectivity* that is meant by the expression. To talk about what is expressed by a given expression may be to talk about any one of these ingredients. (To some extent Husserl avoided the usual sort of technical distinctions among semantical relations, specifically rejecting Frege’s sense-reference distinction as a violation of the ordinary use of the words *Sinn* and *Bedeutung*, words which Husserl used interchangeably [2.53].)

Somewhat more precisely, an expression used in ordinary circumstances for purposes of communication may be described as “manifesting” the psychical experience of its producer—that is, the meaning-endowing act—which is a necessary condition of its status as an expression. This manifesting function of an expression would, however, be lacking in the case of an expression used in an unoverheard monologue. The manifesting and the more strictly expressing functions differ also in that, for example, the expression “the three altitudes of a triangle intersect in a point” manifests a distinct mental state or act each time it is used in ordinary circumstances for purposes of communication, although what it expresses, in the stricter sense, remains the same on all occasions of such use.

Some of Husserl’s main points in “Expression and Meaning” are summarized in sections dealing with “equivocations” associated with discussions of meaning and meaninglessness (2.52–61). “A meaningless expression is, properly speaking, not an expression at all.” Thus “green is or” (Husserl’s example) only gives the appearance of an expression (2.54). Meaningfulness, however, entails reference (*Beziehung*) to an object, regardless of whether that object exists or is “fictive.” “Consequently, to use an expression with sense and to refer to an object (to present an object) are one and the same.” Nevertheless, Husserl was careful to point out, the object of an expression is not to be confused with its meaning (2.54). As a result, “objectlessness” of an expression is not “meaninglessness” (where “objectlessness” indicates only the lack of a real object). Neither the name “golden mountain” nor

the name “round square” is meaningless, although both are objectless, the second one necessarily so (2.55). After a rather obscure passage (2.56–57) in which Husserl was evidently criticizing (without mentioning) pragmatism for identifying meaning with meaning-fulfilling acts, he devoted an entire section to the criticism of Mill’s doctrine of connotation and denotation, with particular attention to Mill’s view of “non-connotative names” as meaningless. A proper name, Husserl objected, is not a mere sign but an expression. It can, like any other expression, function as a mere sign—for instance, in a signature—but it ordinarily does much more. He felt that if Mill’s distinction between what a name denotes and what it connotes were carefully separated from the merely related distinction between what a name names and what it means, some of the confusion in Mill’s doctrine would be dissipated (2.57–61).

In his fourth Logical Investigation, “The Distinction of Independent and Dependent Meanings and the Idea of a Pure Grammar” (2.294–342), Husserl pursued the analysis of meaning undertaken in the earlier treatise. Of most historical interest is his attempted refurbishing of the Enlightenment project of a universal grammar, an enterprise furthered by Anton Marty (1847–1914), another of Brentano’s students, in his *Grundlegung der allgemeinen Grammatik und Sprachphilosophie* (1908). In his treatise Husserl developed a notion of “pure logic” as “the pure formal theory of *meanings*” and insisted that we could not understand the functioning of even our own language if we did not first construct a “pure-logical grammar,” the subject matter of which would be the “ideal form” of language. At a later stage of his career, however, Husserl abandoned this “ideal-language” approach to considerations of semantics and syntax and urged the return to living history and actual speech—the return to the *Lebenswelt*—for the materials of philosophy.

Husserl’s influence in all respects has been felt more strongly in Europe than in England and America. Some of his work in philosophy of language has been investigated and developed further by, among others, Maurice Merleau-Ponty.

Alexius Meinong (1853–1920) developed his theory of linguistic meaning as an integral part of the “theory of objects” in which he worked out his version of the doctrine of intentionality. His most complete presentation of it may be found in his *Ueber Annahmen* (1902; rev. ed., 1910).

Meinong began, in the traditional way, by developing a semantics of words. His assimilation of it to his theory of objects gave rise to no particularly novel features.

“Whoever happens to pronounce the word ‘sun,’ he declared, “normally gives *expression* [*Ausdruck*] thereby, whether or not he wishes to do so, that a definite presentation—it may be a presentation of perception or one of imagination—is taking place within him. What kind of presentation it is is determined principally on the basis of what is presented in it—i.e., its object—and this object is precisely that which the word ‘sun’ *refers to* [*bedeutet*]” (*Ueber Annahmen*, pp. 19–20; italics added). He summed up his account of the “expression” and “reference” of words (which he presented explicitly as opposed to Husserl’s doctrine of *Ausdruck* and *Bedeutung*) by saying that “a word always ‘refers to’ the object of the presentation that it ‘expresses’ and, conversely, expresses the presentation of the object that it refers to” (p. 20). (The obvious similarities to Frege’s doctrine of sense and reference, even as to the same unusual use of *bedeuten*, may be coincidental. There are, of course, clear differences as well, especially as regards Frege’s treatment of “sense” and Meinong’s treatment of “expression.”) Meinong concluded his rather brief account of the meaning of words by refining his original distinction to the point of recognizing a “secondary” as well as the “primary” expression and reference described above (pp. 20–23).

He then undertook to apply his “antithesis of expression and reference” to the semantics of sentences, and he first applied it in an effort to provide a more satisfactory criterion of sentencehood than that provided by traditional grammar. The phrase “the blue sky” and the sentence “the sky is blue” have, he maintained, one and the same object as their reference. If, however, “I say ‘the sky is blue,’ I thereby *express* an opinion [*Meinung*], a judgment, that can in no way be gathered from the words ‘the blue sky’” (p. 25; italics added). The phrase expresses the kind of experience described by Meinong as the pure presentation or idea, the *Vorstellung* proper, while the sentence expresses a different fundamental kind of experience, the judgment (*Urteil*). The judgment differs from the pure presentation giving rise to it in two respects that might be described as “intentional”—conviction and a determinate position as regards affirmation and negation (p. 2). Sentences, he claimed, might also be used to express “assumptions” (*Annahmen*), which, because they have to be either affirmative or negative assumptions, share the second defining characteristic of judgments but lack the first, conviction (p. 4).

Meinong’s most important contribution to semantics, partly because of its effect on the development of Russell’s theory of descriptions, was his doctrine of “objectives,” particularly in his application of it in the

treatment of negative sentences, which he recognized to be crucial cases for his doctrine. Suppose that a magistrate judges that on a given occasion there was no disturbance of the peace. On the Brentano-Meinong view of mental states, there must be an object of that judgment. Putting it another way, there must be a reference for the sentence “there has been no disturbance of the peace,” which expresses the magistrate’s judgment. It cannot be the disturbance of the peace on the occasion in question, for, by the hypothesis, there is no such object. According to Meinong it can, however, be “the non-existence of a disturbance of the peace” or “that there has been no disturbance of the peace.” Meinong held that it makes no sense to say that that nonexistence *exists*, but we may say that “it is the case.” This entity, the being of which is being the case, is the “objective” to which the sentence refers. The objective *may* be a fact—if, for example, it is a fact that there has been no disturbance of the peace—but false judgments also have their objectives (2nd ed., p. 43). That regarding which the judgment is made—a disturbance of the peace—is the object (proper) of the judgment; what is judged in it—that there has been no disturbance of the peace—is its objective (2nd ed., p. 52).

The objectives of negative sentences and the objects of denials of existence, such as “a *perpetuum mobile* does not exist,” “must have properties, and even characteristic properties, for without such the belief in non-existence can have neither sense nor justification; but the possession of properties is as much as to say a manner of being [*Sosein*],” which of course is not to be confused with existence. “In this sense ‘there are’ also objects that do not exist, and I have expressed this in a phrase that, while rather barbarous, I am afraid, is hard to improve upon—viz. ‘externality [*Aussersein*] of the pure object’” (p. 79). Meinong believed that he had formulated an important principle in this doctrine, “the principle of the independence of manner of being from existence,” which he illustrated and summarized in the following famous passage: “Not only is the often cited golden mountain golden, but the round square, too, is as surely round as it is square.... To know that there are no round squares, I have to pass judgment on the round square.... Those who like paradoxical expressions can therefore say: there are objects of which it is true that there are no objects of that kind” (*Ueber Gegenstandstheorie*, 1904, pp. 7ff.). Meinong’s influence on the development of semantics is best exhibited in Russell’s series of articles on him in *Mind*, 1899–1907.

**See also** Semantics, History of [Addendum].



## Bibliography

There is at present no thorough general study of the history of semantics. Much relevant information may be found, however, in works on the history of logic, particularly in I. M. Bocheński, *A History of Formal Logic*, translated by Ivo Thomas (Notre Dame, IN, 1961); Philotheus Boehner, *Medieval Logic* (Manchester, U.K.: Manchester University Press, 1952); L. M. de Rijk, *Logica Modernorum*, Vol. I (Assen: Van Gorcum, 1962); William and Martha Kneale, *The Development of Logic* (Oxford: Clarendon Press, 1962); Benson Mates, *Stoic Logic* (Berkeley: University of California Press, 1953); E. A. Moody, *The Logic of William of Ockham* (New York and London: Sheed and Ward, 1935), and *Truth and Consequence in Mediaeval Logic* (Amsterdam: North-Holland, 1953); and Carl Prantl, *Geschichte der Logik im Abendlande*, 4 vols. (Leipzig: Hirzel, 1855–1870).

Surveys less thorough than this entry are contained in Ch. 7, "Language," of Janet and Séailles's *History of the Problems of Philosophy*, translated by A. Monahan and edited by H. Jones (New York: Macmillan, 1902), and in Ch. 1, "The Problem of Language in the History of Philosophy," of Vol. I, *Language*, of Ernst Cassirer's *The Philosophy of Symbolic Forms*, translated by R. Manheim (New Haven, CT: Yale University Press, 1953).

The following selection of books and articles is arranged in four parts corresponding to the four parts of the article.

### ANTIQUITY

Allen, W. S. "Ancient Ideas on the Origin and Development of Language." *Transactions of the Philosophical Society* [of Great Britain] (1948): 35–60.

Billiesich, F. *Epikurs Sprachphilosophie*. Landskron, 1912.

De Lacy, Estelle. "Meaning and Methodology in Hellenistic Philosophy." *Philosophical Review* 47 (1938): 390–409.

De Lacy, Phillip. "The Epicurean Analysis of Language." *American Journal of Philology* 60 (1939): 85–92.

Demos, Raphael. "Plato's Philosophy of Language." *Journal of Philosophy* 61 (1964): 595–610.

Deuschle, Julius. *Die Platonische Sprachphilosophie*. Marburg, 1852.

Goldschmidt, Victor. *Essai sur le "Cratyle"*. Paris: Champion, 1940.

Lersch, Laurenz. *Die Sprachphilosophie der Alten*. 3 vols. Bonn: H.B. König, 1838–1841.

Luce, J. V. "The Theory of Ideas in Plato's *Cratylus*." *Phronesis* 10 (1965): 21–36.

Moravcsik, Julius. "Being and Meaning in the *Sophist*." *Acta Philosophica Fennica*, Fasc. 14 (1962): 23–78.

Robinson, Richard. "A Criticism of Plato's *Cratylus*." *Philosophical Review* 65 (1956): 324–341.

Robinson, Richard. "The Theory of Names in Plato's *Cratylus*." *Revue internationale de philosophie* 9 (1955): 221–236.

Schoemann, G. F. *Die Lehre von den Redetheilen nach den Alten*. Berlin, 1862.

Steinthal, Hugo. *Die Geschichte der Sprachwissenschaft bei den Griechen und Römern*. Berlin: F. Dümmlers, 1863.

Weltring, G. *Das Σημεῖον in der Aristotelischen, stoischen, Epikureischen, und skeptischen Philosophie*. Bonn, 1910.

### THE MIDDLE AGES

Arnold, Erwin. "Zur Geschichte der Suppositionstheorie, die Wurzeln des modernen Subjektivismus." *Symposion, Jahrbuch für Philosophie* 3 (1952): 1–34.

Boehner, Philotheus. "A Medieval Theory of Supposition." *Franciscan Studies* 18 (1958): 240–289.

Boehner, Philotheus. "Ockham's Theory of Signification." *Franciscan Studies* 6 (1946): 143–170.

Boehner, Philotheus. "Ockham's Theory of Supposition and the Notion of Truth." *Franciscan Studies* 6 (1946): 261–292.

Boehner, Philotheus. "Ockham's Theory of Truth." *Franciscan Studies* 5 (1945): 138–161.

Carreras y Artau, Joaquin. *De Ramón Lull á los modernos ensayos de formación de una lengua universal*. Barcelona: Consejo Superior de Investigaciones Científicas, Delegación de Barcelona, Instituto Antonio de Nebrija, 1946.

Grabmann, Martin. "Die Entwicklung der mittelalterlichen Sprachlogik (Tractatus de Modis Significandi)." In his *Mittelalterliches Geistesleben*. 3 vols. Munich: Hueber, 1926. Vol. I, Ch. 4.

Grabmann, Martin. *Thomas von Erfurt und die Sprachlogik des mittelalterlichen Aristotelismus; Sitzungsberichte der Bayerischen Akademie der Wissenschaften* (Philosophisch-historische Abteilung), Jahrgang 1943, Heft 2.

Heidegger, Martin. *Die Kategorien- und Bedeutungslehre des Duns Scotus*. Tübingen: Mohr, 1916.

Henry, D. P. "Saint Anselm's Nonsense." *Mind* 72 (1963): 51–61.

Manthey, Franz. *Die Sprachphilosophie des hl. Thomas von Aquin und ihre Anwendung auf Probleme der Theologie*. Paderborn, Germany, 1937.

Prior, A. N. "The Parva Logicalia in Modern Dress." *Dominican Studies* 5 (1952): 78–87.

Roberts, Louise Nisbet. "Supposition: A Modern Application." *Journal of Philosophy* 57 (1960): 173–182.

Rotta, P. *La filosofia del linguaggio nella patristica e nella scolastica*. Turin, 1909.

Saw, Ruth Lydia. "William of Ockham on Terms, Propositions, Meaning." *PAS* 42 (1941–1942): 45–64.

Thomas, Ivo. "Saint Vincent Ferrer's *De Suppositionibus*." *Dominican Studies* 5 (1952): 88–102.

Thurot, Charles. *Notices et extraits des manuscrits latins pour servir à l'histoire des doctrines grammaticales au moyen âge*. Vol. 22, Part 2 of *Notices et extraits des manuscrits de la Bibliothèque Impériale*. Paris, 1868.

Wallerand, G. "Étude sur Siger de Courtrai." In *Les oeuvres de Siger de Courtrai (Étude critique et textes inédits)*. Vol. VIII of *Les philosophes belges*. Louvain: Institut supérieur de philosophie de l'université, 1913.

### RENAISSANCE AND ENLIGHTENMENT

Acton, H. B. "The Philosophy of Language in Revolutionary France." *Proceedings of the British Academy* 45 (1959): 199–219.

Cohen, L. Jonathan. "On the Project of a Universal Character." *Mind* 63 (1954): 49–63.

François, Alexis. "La Grammaire philosophique." In *Histoire de la langue française des origines à 1900*, by F. Brunot. Paris: A. Colin, 1932. Vol. VI, Part 2, Book 2.

Friedrich, Hugo. "Die Sprachtheorie der französischen Illuminaten des 18. Jahrhunderts, insbesondere Saint-

- Martins." *Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte* 13 (1935): 293–310.
- Funke, Otto. *Studien zur Geschichte der Sprachphilosophie: I. Zur Sprachphilosophie des 18. Jahrhunderts: J. Harris' "Hermes"; II. Zur Sprachphilosophie der Gegenwart*. Bern: A. Francke, 1927.
- Funke, Otto. *Zum Weltsprachenproblem in England im 17. Jahrhundert: G. Dalgarno's "Ars signorum" (1661) und J. Wilkins' "Essay towards a Real Character and a Philosophical Language."* *Anglistische Forschungen*, no. 69. Heidelberg: C. Winter, 1929.
- Furlong, E. J. "Berkeley's Theory of Meaning." *Mind* 73 (1964): 437–438.
- Harnois, Guy. *Les théories du langage en France de 1660 à 1821*. Paris: Société d'édition "Les belles lettres," 1929.
- Hunt, H. J. "Logic and Linguistics: Diderot as 'Grammairien-philosophe.'" *Modern Language Review* 33 (1938): 215–233.
- Kuehner, Paul. *Theories on the Origin and Formation of Language in the Eighteenth Century in France*. Philadelphia: n.p., 1944.
- Maynial, Édouard. "Les grammairiens-philosophes du XVIII<sup>e</sup> siècle: La Grammaire de Condillac." *Revue Bleue*, 4th series, 19 (1903): 317–320.
- Oesterle, John A. "Another Approach to the Problem of Meaning." *Thomist* 7 (1944): 233–263. Supposition theory in John of St. Thomas.
- O'Flaherty, James C. *Unity and Language: A Study in the Philosophy of Johann Georg Hamann*. Chapel Hill: University of North Carolina, 1952.
- Pott, August. "Zur Geschichte und Kritik der sogenannten allgemeinen Grammatik." *Zeitschrift für Philosophie und Philosophische Kritik* 43 (1863): 102–141.
- Robert, Louis. *Les théories logiques de Condillac*. Paris, 1869.
- Russell, L. J. "Note on the Term Σημειωτική in Locke." *Mind* 48 (1939): 405–406.
- Sahlin, G. *César Chesneau du Marsais et son rôle dans l'évolution de la grammaire générale*. Paris: Presses Universitaires de France, 1928.
- Salvucci, Pasquale. *Linguaggio e mondo umano in Condillac*. Urbino, Italy, 1957.
- Trendelenburg, Adolf. "Ueber Leibnizens Entwurf einer allgemeinen Charakteristik." In his *Historische Beiträge zur Philosophie*, Vol. III. Berlin, 1867.
- Unger, Rudolf. *Hamanns Sprachtheorie im Zusammenhang seines Denkens*. Munich, 1905.
- Wallace, Karl R. *Francis Bacon on Communication and Rhetoric*. Chapel Hill: University of North Carolina Press, 1943.
- THE NINETEENTH AND TWENTIETH CENTURIES
- Bronstein, Daniel. "Inquiry and Meaning." In *Studies in the Philosophy of Charles Sanders Peirce*, edited by P. P. Wiener and F. H. Young, 33–52. Cambridge, MA: Harvard University Press, 1952.
- Buchler, Justus. "What Is the Pragmaticist Theory of Meaning?" In *Studies in the Philosophy of Charles Sanders Peirce*, edited by P. P. Wiener and F. H. Young, 21–32. Cambridge, MA: Harvard University Press, 1952.
- Burks, Arthur W., and Paul Weiss. "Peirce's Sixty-six Signs." *Journal of Philosophy* 42 (1945): 383–388.
- Dewey, John. "Peirce's Theory of Linguistic Signs, Thought, and Meaning." *Journal of Philosophy* 43 (1946): 85–95.
- Farber, Marvin. *The Foundation of Phenomenology: Edmund Husserl and the Quest for a Rigorous Science of Philosophy*. Cambridge, MA: Harvard University Press, 1943.
- Findlay, J. N. *Meinong's Theory of Objects*. Oxford: Oxford University Press, H. Milford, 1933.
- Lovejoy, Arthur O. "What Is the Pragmaticist Theory of Meaning? The First Phase." In *Studies in the Philosophy of Charles Sanders Peirce*, edited by P. P. Wiener and F. H. Young, 3–20. Cambridge, MA: Harvard University Press, 1952.
- Mohanty, J. N. *Edmund Husserl's Theory of Meaning*. The Hague: M. Nijhoff, 1964.
- Morris, Charles W. "Signs about Signs about Signs." *Philosophy and Phenomenological Research* 9 (1948–1949): 115–133.
- Ogden, C. K. *Bentham's Theory of Fictions*. London: K. Paul, Trench, Trubner, 1932.
- Ogden, C. K., and I. A. Richards. "Some Moderns—1. Husserl, 2. Russell, 3. Frege, 4. Gomperz, 5. Baldwin, 6. Peirce." In *The Meaning of Meaning*. New York: Harcourt Brace, 1923. Appendix D.
- Rynin, David. "Introduction" and "Critical Essay." In his edition of Alexander Bryan Johnson's *A Treatise on Language*. Berkeley and Los Angeles, 1947.
- Walker, Jeremy D. B. *A Study of Frege*. Ithaca, NY: Cornell University Press, 1965.
- Weiler, Gershon. "On Fritz Mauthner's Critique of Language." *Mind* 67 (1958): 80–87
- Wienpahl, P. D. "Frege's Sinn und Bedeutung." *Mind* 59 (1950): 483–494.
- The development of philosophy of language, like the development of philosophy generally in the twentieth century, has taken place at least as much in articles as in books, and the use of anthologies is consequently almost indispensable in the study of it. All the following anthologies contain material useful for the study of philosophy of language in the twentieth century: *Logical Positivism*, edited by A. J. Ayer (Glencoe, IL: Free Press, 1959); *Philosophical Analysis*, edited by Max Black (Ithaca, NY: Cornell University Press, 1950); *Philosophy and Ordinary Language*, edited by C. E. Caton (Urbana: University of Illinois Press, 1963); *Ordinary Language*, edited by V. C. Chappell (Englewood Cliffs, NJ: Prentice-Hall, 1964); *Logic and Language*, edited by A. G. N. Flew (1st series, New York, 1951; 2nd series, New York: Philosophical Library, 1953); *The Structure of Language*, edited by J. Fodor and J. Katz (Englewood Cliffs, NJ: Prentice-Hall, 1964); *Classics in Semantics*, edited by D. E. Hayden and E. P. Alworth (New York: Philosophical Library, 1965); and *Semantics and the Philosophy of Language*, edited by L. Linsky (Urbana: University of Illinois Press, 1952).
- William P. Alston, *Philosophy of Language* (Englewood Cliffs, NJ: Prentice-Hall, 1964), provides a good introduction to the subject as it developed in the first half of the twentieth century. John Passmore, *A Hundred Years of Philosophy* (London: Duckworth, 1957), provides an excellent history of this period and its immediate background, with considerable detail regarding developments in philosophy of language. More specialized and also to be recommended are G. J. Warnock, *English Philosophy since 1900* (New York and

London: Oxford University Press, 1958), and J. O. Urmson, *Philosophical Analysis: Its Development between the Two World Wars* (London: Clarendon Press, 1956).

*Norman Kretzmann (1967)*

## SEMANTICS, HISTORY OF [ADDENDUM]

In the 1960s, the semantics in vogue in linguistics seems to have favored some kind of decompositional approach. Consider kinship terms. Taking *P* to mean “parent of” and *F* to mean *female* you can analyze most kinship terms using first-order predicate logic. So “*x* is *y*’s aunt” would come out as  $\exists z \exists w (Pwy \ \& \ Pzw \ \& \ Pzx \ \& \ w \neq x \ \& \ Fx)$ . Some linguists, notably George Lakoff (1971) and James McCawley (1972), championed what was called generative semantics, where the idea was that the base level was a semantic level of structures in something such as first-order logic, which could be converted to a surface level. Noam Chomsky’s (1965) level of “deep structure” was thought to lie somewhere between what subsequently became known as the level of “logical form” and the surface level. There were then debates about how autonomous syntax is, and how much of it depends on semantic input. Chomsky himself in *Aspects of the Theory of Syntax* gave an important place to what were called selectional restrictions, one of whose jobs was to rule out sentences such as “Colorless green ideas sleep furiously.”

None of this work however addressed the philosophical question of what semantics really is. For that, the input came from logic. At least since Ludwig Wittgenstein’s *Tractatus Logico-Philosophicus* (1963) and Rudolf Carnap’s *Meaning and Necessity* (1947), the idea had been around that the clue to semantic understanding is in the notion of the truth conditions of a sentence. How might you tell the difference between someone who knows English and someone who does not? You point to a table and utter the sentence: “There is a pen on that table.” First you utter it when the pen is on the table, and then you utter it when the pen has been taken away. You do not need to know English to know the difference between a situation in which there is a pen, and a situation in which there is not, but you do need to know English to know that that difference is correlated with the truth or falsity of a sentence of English. This insight had been at the heart of the languages of logic.

In particular, during the 1960s there emerged the possible worlds semantics for modal logic, whereby an interpretation to a logic provides a set of possible worlds,

and each wff of the formal language has a truth value in each of these worlds. In modal logic the necessity operator written *L* or  $\Box$  is so interpreted that  $\Box\alpha$  is true in a world *w* iff  $\alpha$  is true in every *w'* that is possible relative to *w*, where relative possibility is specified by a relation of accessibility between worlds. The truth at a world of every complex wff is determined by the semantic evaluation rules associated with each way of getting a more complex wff from its simpler parts—rules such as the one for  $\Box$  just mentioned and the rule that  $\sim\alpha$  is true in a world iff  $\alpha$  is false in that world, and so on. A. N. Prior was at the same time developing interpretations that used moments of time as the semantical indices and interpreted accessibility as the temporal ordering. Such logics were developed for epistemic and deontic operators, though there were those who, mostly under the influence of W. V. Quine (1953), declared that none of this was genuine logic.

But at that time there was still a rift between the logicians, who argued that you could only discuss philosophical issues precisely and profitably in a language that was free from the vagueness and imprecision of natural language, and the ordinary language philosophers who argued that philosophical problems come up in ordinary language and that, as P. F. Strawson had famously claimed, “ordinary language has no exact logic.” The key figure at this point was Richard Montague. During the 1960s Montague began to realize that the languages of intensional logic (logics that involved truth at indices such as worlds, times, and the like) could be used in the semantical study of natural language. Noting that, under Chomsky’s influence, linguists were beginning to realize that the grammatical structures of English sentences could be produced by a set of formal rules, Montague was able to produce fragments of English in which the syntactical rules, such as subject-predicate combination, could be assigned semantic interpretations, which specified how the truth-conditional meaning of a complex structure derived by these syntactical rules could be obtained from the meanings of the simpler parts, from which the complex structure is constructed. Montague’s work (collected in his *Formal Philosophy*) was taken up by Barbara Partee and her students first at UCLA and then at the University of Massachusetts, and has become the dominant tradition in semantics. Montague’s work established a revolution in philosophy as well as linguistics because it became no longer tenable to maintain the distinction between formal logic and natural language.

An alternative tradition, which also developed during the late 1960s and the 1970s, was inspired by Donald

Davidson's *Truth and Meaning* (1967a). Davidson's insight was to apply Alfred Tarski's (1956) work on the semantics of the ordinary predicate calculus to natural language. Tarski had set the goal of semantics as the generation of what came to be called T-schemata. The best known example of a T-schema is "'Snow is white' is true iff snow is white," where the left-hand side says that a certain sentence is true, and the right-hand side states the conditions under which it is true by using the sentence itself. While this has an air of triviality, it is not trivial. The T-schema does not say that "Snow is white" is true iff it is true. It says that it is true iff snow is white, and in a language in which the words are used differently, "Snow is white" would not be true iff snow is white. Suppose that in English *snow* means what *grass* means in English, and vice versa, and that *white* means what *green* means in English. Then the corresponding T-schema in the new language would now be "'Snow is white' is true iff snow is white," but it would now say that "'Snow is white' is true iff grass is green."

Davidson's theory does not make use of worlds or times, and he uses various tricks to get their effect. His theory of indirect discourse (*On Saying That*) treats the *that* in "Galileo said that the earth moves" as being a demonstrative pronoun whose referent is a prior utterance of "The Earth moves," and where the utterer claims that that utterance make him and Galileo samesayers. Another area where the Davidson approach has been applied is that of adverbial modification. Sentential adverbs such as *possibly* or *unfortunately* are the kind that are studied by semantics in terms of worlds, times, and other such indices, but words such as *competently* also need analysis because even if those who drive and those who sing are the same people, those who drive competently need not be the same people as those who sing competently. Davidson's account of adverbial modification (*The Logical Form of Action Sentences*) treats events as individuals, and would analyze, say, "Alice sings competently" as  $\exists x(x$  is the singing of a song by Alice and  $x$  is competent). Davidson's use of events has been extended by Terry Parsons (1990).

The development of formal semantics raised questions about the connection between the semantical theories provided by logic and the practice of linguistic communication. One way of connecting language with communicative practice is via propositional attitudes, of which the principal ones are knowledge and belief. David Lewis (in *Languages and Language* [1975]) defines a formal model as the correct one for a population if they only produce sentences that are true according to the model

when they have certain beliefs, and trust others to do the same. And Robert Stalnaker (1984) engages with Hartry Field (1978) on the question of whether meaning depends on beliefs, or whether beliefs can only be defined in terms of an internal language.

The connection between meaning and belief forms the focus of one of the most widely discussed issues in the philosophy of language, mainly because of the views of two influential philosophers, Saul Kripke and Hilary Putnam. Kripke (1972) talks about the fact that we can use the word *Aristotle* even though we may know almost nothing about him. Kripke also imagines a person named *Pierre* (A puzzle about belief) who has heard about a beautiful city called *Londres*, which he believes is pretty. But the part of London that he comes to live in is so dismal that he believes that the city he knows as London is not at all pretty. So, Kripke asks us, does Pierre believe that London is pretty? Putnam (1975) looks at our use of the word *water*. If this word means  $H_2O$ , then it would seem that those without a background in chemistry do not know the meaning of what they say.

Recent developments are too many and various to detail adequately here. One, for instance, has involved the claim that semantics should be dynamic. By this, it is meant that the meaning of a sentence should be thought of not so much in terms of its truth conditions as in terms of the potential to change truth conditions. Advocates of dynamic semantics tackle this by thinking of the utterance of a sentence as like a computer program. Instead of a sentence being true or false at an index, think of the first index as being how things are before the sentence is uttered and a second index as being how things are after the sentence has been uttered. One version of this thought appears in *Dynamic Predicate Logic*, found in Jeroen Groenendijk's and Martin Stokhof's 1991 paper. Other dynamic frameworks include those based on Hans Kamp's *Discourse Representation Theory* (or in the more or less equivalent writings of Irene Heim [1983]). A rather different version of a dynamic approach is found in the game-theoretical semantics developed by Jaakko Hintikka (1983).

If the meaning of a sentence is the set of indices at which it is true, we have the consequence that all sentences true at the same set of indices have the same meaning. So for instance, because all mathematical truths are true in all possible worlds, mathematical knowledge would be trivial. However, although the entities used in formal semantics have the job of delivering a set of indices as the final result, any interpreted sentence generates a semantic structure made up from the semantic val-

ues of its simple parts (Lewis 1972, Cresswell 1985). This structure can then provide the input to propositional attitude operators. The kind of indices that have been used in classical truth-conditional semantics are complete in the sense of deciding every sentence—or at least of every sentence that might be said to have a truth value at all. In the 1980s Jon Barwise and John Perry's *Situation Semantics* was based on the view that the entities used in semantics should not be complete. The meaning of a sentence such as "Sebastian laughs" would not be the set of indices at which Sebastian laughs, but would be referred to as a situation, which would be composed of Sebastian, a location, and the relation of laughing that holds between Sebastian and the location at which he laughs.

**See also** Semantics.

### Bibliography

- Bach, E. *Informal Lectures on Formal Semantics*. Albany: State University of New York Press, 1989.
- Barwise, J., and J. Perry. *Situations and Attitudes*. Cambridge MA: MIT Press, 1983.
- Benthem, J. F. A. K. van, and A. ter Meulen, eds. *Handbook of Logic and Language*. Amsterdam: Elsevier/MIT Press, 1997.
- Cann, R. *Formal Semantics*. Cambridge, U.K.: Cambridge University Press, 1993.
- Carnap, R. *Meaning and Necessity*. Chicago: University of Chicago Press, 1947.
- Chierchia, G., and S. McConnell-Ginet. *Meaning and Grammar: An Introduction to Semantics*. Cambridge, MA: MIT Press, 1990.
- Chomsky, N. *Aspects of The Theory of Syntax*. Cambridge, MA: MIT Press, 1965.
- Cresswell, M. J. *Logics and Languages*. London: Methuen, 1973.
- Cresswell, M. J. "Semantic Competence." In *Meaning and Translation*, edited by F. Guenther and M. Guenther-Reutter. London: Duckworth, 1978.
- Cresswell, M. J. *Structured Meanings*. Cambridge, MA: MIT Press, 1985.
- Davidson, D. "The Logical Form of Action Sentences." In *The Logic of Decision and Action*, edited by N. Rescher. Pittsburgh, PA: University of Pittsburgh Press, 1967b.
- Davidson, D. "Truth and Meaning." *Synthese* 17 (1967a): 304–323.
- Davidson, D. "On Saying That." In *Words and Objections: Essays on the Work of W. V. Quine*. Dordrecht: Reidel, 1969.
- Davidson D., and G. Harman, eds. *Semantics of Natural Language*. Dordrecht: Reidel, 1972.
- Dowty, D. R., R. Wall, and P. S. Peters. *Introduction to Montague Semantics*. Dordrecht: Reidel, 1981.
- Field, H. "Mental Representation." *Erkenntnis* 13 (1978): 9–61.
- Gamut, L. T. F. *Logic, Language, and Meaning*. Chicago: University of Chicago Press, 1991.
- Groenendijk, J., and M. Stokhof. "Dynamic Predicate Logic." *Linguistics and Philosophy* 14 (1991): 39–100.
- Heim, I. "File Change Semantics and the Familiarity Theory of Definiteness." In *Meaning, Use, and Interpretation of Language*, edited by R. Bäuerle. Berlin: de Gruyter, 1983.
- Heim, I., and A. Kratzer. *Semantics in Generative Grammar*. Malden, MA: Blackwell, 1998.
- Hintikka, K. J. J. *Anaphora and Definite Descriptions: Two Applications of Game-Theoretical Semantics*. Dordrecht: Reidel, 1985.
- Hintikka, K. J. J., and J. Kulas. *The Game of Language*. Dordrecht: Reidel, 1983.
- Kamp, J. A. W. "A Theory of Truth and Semantic Representation." In *Formal Methods in The Study of Grammar*, edited by J. A. G. Groenendijk. Amsterdam: Mathematische Centrum, 1983.
- Keenan, E. L., ed. *Formal Semantics of Natural Language*. Cambridge, U.K.: Cambridge University Press, 1975.
- Kripke, S. A. "Naming and Necessity." In *Semantics of Natural Language*, edited by D. Davidson and G. Harman. Dordrecht: Reidel, 1972.
- Kripke, S. A. "A Puzzle about Belief." In *Meaning and Use*, edited by A. Margalit Reidel: Dordrecht, 1979.
- Lakoff, G. "On Generative Semantics." In *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology*, edited by D. D. Steinberg and L. A. Jakobovits. Cambridge, U.K.: Cambridge University Press, 1971.
- Lappin, S., ed. *The Handbook of Contemporary Semantic Theory*. Oxford: Blackwell, 1996.
- Lewis, D. K. *Counterfactuals*. Oxford: Basil Blackwell, 1973.
- Lewis, D. K. "General Semantics." In *Semantics of Natural Language*, edited by D. Davidson and G. Harman. Dordrecht: Reidel, 1972.
- Lewis, D. K. "Languages and Language." In *Language Mind and Knowledge*, edited by K. Gunderson. Minneapolis: University of Minnesota Press, 1975.
- McCawley, J. D. "A Program for Logic." In *Semantics of Natural Language*, edited by D. Davidson and G. Harman. Dordrecht: Reidel, 1972.
- Montague, R. M. *Formal Philosophy*. New Haven, CT: Yale University Press, 1974.
- Parsons, T. *Events in The Semantics of English*. Cambridge, MA: MIT Press, 1990.
- Prior, A. N. *Past, Present, and Future*. Oxford: Clarendon Press, 1967.
- Prior, A. N. *Time and Modality*. Oxford: Clarendon Press, 1957.
- Putnam, H. "The Meaning of 'Meaning.'" In *Language, Mind, and Knowledge*, edited by K. Gunderson. Minneapolis: University of Minnesota Press, 1975.
- Quine, W. V. O. *A Logical Point of View; 9 Logico-Philosophical Essays*. Cambridge, MA: Harvard University Press, 1953.
- Richard, M. *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them*. Cambridge, U.K.: Cambridge University Press, 1990.
- Stalnaker, R. C. *Inquiry*. Cambridge, MA: MIT Press, 1984.
- Stalnaker, R. C. "A Theory of Conditionals." In *Studies in Logical Theory*, edited by N. Rescher. Oxford: Basil Blackwell, 1968.
- Steinberg, D.D. and L.A. Jakobovits. *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology*. Cambridge, U.K.: Cambridge University Press, 1971.

Tarski, A. "The Concept of Truth in Formalized Languages." In *Logic, Semantics, and Metamathematics*. Oxford: Clarendon Press, 1956.

Wittgenstein, L. *Philosophical Investigations*. Oxford: Blackwell, 1953.

Wittgenstein, L. *Tractatus Logico-Philosophicus*. 2nd ed. Translated by D. F. Pears and B. F. McGinness. London: Routledge and Kegan Paul, 1963.

*M. J. Cresswell (2005)*

## SEN, AMARTYA K. (1933–)

Amartya K. Sen, an economist and philosopher, was born in Bengal in 1933. The memory of the Bengal famine of 1943, in which more than 2 million people died, drew him to work on economics and ethics. He studied economics at Presidency College, Calcutta, and Trinity College, Cambridge, where he received a doctorate in economics in 1959. After he taught at the Delhi School of Economics, the London School of Economics, he held the posts of Drummond Professor of Political Economy at Oxford University (also Fellow of All Souls College), and then Lamont University Professor and professor of economics and philosophy at Harvard University. His contributions to economics lie in the areas of social choice theory, theory of choice, development economics, labor economics, cost-benefit analysis, and the measurement of inequality and poverty. In 1998 he was awarded the Nobel Memorial Prize in Economics "for his contributions to welfare economics" and appointed Master of Trinity College, Cambridge.

### INFORMATIONAL PARSIMONY OF SOCIAL WELFARE JUDGMENT

The large number of Sen's works in economics and philosophy are marked by tireless criticism of utilitarianism and the utilitarian foundations of welfare economics. According to Sen utilitarianism can be factored into three elements: act consequentialism (the goodness of an act is given by the goodness of its consequent states of affairs), welfarism (the goodness of a state of affairs is given by the goodness of utility information regarding that state), and sum-ranking (the goodness of utility information is given by the sum total of different people's utilities). These elements impose informational constraints on policy judgments and economic evaluation: Act consequentialism does not consider the intrinsic value of an act or the motivation underlying the act; welfarism rules out nonutility information such as violation of rights from influencing the goodness of an act; and sum-ranking excludes

information about the state of people who are worse off. Sen holds that the informational basis for judgments of goodness should include nonutility information and information about the distribution of utility among different people.

Much of his philosophical standpoint originates in the close examination of Kenneth J. Arrow's (1921–) general impossibility theorem: there exists no collective decision-making rule that satisfies some seemingly uncontroversial axioms (unrestricted domain, weak Pareto principle, independence of irrelevant alternatives, and nondictatorship). In *Collective Choice and Social Welfare* (1970a) he scrutinizes the formal and philosophical reach of this theorem and points out the informational parsimony of Arrow's framework. Since Abraham Bergson (1914–2003) and Paul A. Samuelson (1915–) established the "new" welfare economics, individual preference orderings have been assumed to be ordinal and interpersonally incomparable, because there is supposed to be no scientific ground to compare one person's preference satisfaction with another's. Sen shows that if the informational basis is extended to include some kind of interpersonal comparability (e.g., the unit comparability or level comparability), there exist collective decision-making rules, including some egalitarian rules such as *maximin* and its lexicographic extension, *leximin* (as endorsed in John Rawls's difference principle).

### INDIVIDUAL FREEDOM AND THE NOTION OF WELL-BEING

In "The Impossibility of a Paretian Liberal" (1970b) he further shows that two conditions in Arrow's theorem (unrestricted domain and weak Pareto principle) are inconsistent with individuals' minimal liberty. Although the weak Pareto principle (if everyone in the society strictly prefers  $x$  to  $y$ ,  $x$  is socially preferred to  $y$ ) is taken to be uncontroversial in economics, it is sufficient to spread the decisiveness of a certain group over all the pair of alternatives, even if the preference over the pair is a purely personal matter that the society should respect. While the same concern led Robert Nozick to his libertarian side-constraint theory, in "Rights and Agency" (1982), Sen adopts a broadly defined consequentialist theory called a goal-right system, according to which individual freedom should be promoted as an end by the society. This moves him to give individual freedom a central role in the evaluation of states of affairs.

In *Commodities and Capabilities* (1985a) he argues against the "opulence" view of well-being (e.g., real income and Rawls's primary goods) and the "utility" view

(e.g., happiness, desire-fulfillment, and the revealed preference theory in welfare economics), and proposes an alternative notion of well-being: the capability to function. He takes human life to consist in a combination of various doings and beings, which he calls functionings (e.g., moving, being well nourished, being in good health, and being socially respected). The capability to function refers to different combinations of functionings, and the capability of a person corresponds to freedom to choose one kind of life among others.

One advantage of this approach is that it takes account of the people's varying capacities to convert primary goods into abilities to pursue their ends. Each person's capability to function is influenced by internal factors such as disability, illness, age, and gender, as well as external factors such as climatic circumstances, educational arrangement, the prevalence of crime and violence, and the resource distribution within the family. What a disabled person can achieve from a larger set of goods may be much less than what an able-bodied person can achieve from a smaller set of goods. The capability approach offers the analytic ground to capture people's diverse needs. This approach changed not only the concept of well-being in ethics but also the paradigm of international development. It became the source of the Human Development Indicators of the United Nations Development Programme.

Through his empirical studies on famines, in *Poverty and Famines: An Essay on Entitlement and Deprivation* (1981), Sen maintains that there has never been a famine in a functioning multiparty democracy: The democratic poor countries such as India, Botswana, or Zimbabwe managed to avert famines despite serious crop failure, whereas the dictatorial countries had major famines. This is because, Sen claims, democracy would spread the penalty of famine to the ruling parties and political leaders, thus providing the political incentives to try to prevent any threatening famine. In subsequent works he champions the notion of human rights for their intrinsic importance, their consequential role in providing the political incentives for economic security, and their constructive role in the genesis of values and priorities.

**See also** Consequentialism; Philosophy of Economics; Rights; Utilitarianism.

## Bibliography

### WORKS BY SEN

*Collective Choice and Social Welfare*. San Francisco: Holden-Day, 1970a.

- "The Impossibility of a Paretian Liberal." *Journal of Political Economy* 78 (1970b): 152–157.
- On Economic Inequality*. Oxford, U.K.: Clarendon Press, 1973.
- "Equality of What?" In *Tanner Lectures on Human Values*. Vol. 1., edited by S. McMurrin. Salt Lake City: University of Utah Press, 1980.
- Poverty and Famines: An Essay on Entitlement and Deprivation*. Oxford, U.K.: Clarendon Press, 1981.
- "Rights and Agency." *Philosophy and Public Affairs* 11 (1) (1982): 3–39.
- Commodities and Capabilities*. New York: North-Holland, 1985a.
- "Well-being, Agency, and Freedom: The Dewey Lectures 1984." *Journal of Philosophy* 82 (4) (1985b): 169–221.
- Development as Freedom*. New York: Knopf, 1999.

Iwao Hirose (2005)

## SENECA, LUCIUS ANNAEUS (4 BCE–65 CE)

A Roman adherent of Stoicism with a particular interest in ethics, Seneca had an extensive career in politics and literature. His *Moral Epistles*, two major treatises, and a series of essays including *On Anger* offer an engaging presentation of philosophical ideas and are an important source for earlier Stoic thought. Also extant are eight plays and a political lampoon.

Seneca was of provincial origin, having been born at Córdoba in southern Spain, but was brought to Rome at an early age. There he received an extensive education in public speaking and literary composition. His knowledge of philosophy came from the lecturers Papirius Fabianus and Sotion (both adherents of Sextian moral philosophy), from the Cynic Demetrius, and from the Stoic Attalus. He won considerable repute as an advocate, but his health was poor and he was in disfavor with the emperors Gaius and Claudius. Exiled by Claudius to Corsica, he was recalled in 49 to become tutor in rhetoric to the boy Nero.

Following Nero's accession he held a position of considerable influence, restraining the young ruler's excesses and composing important speeches for him; as Miriam Griffin has shown, however, his influence on administrative policy was much less than accounts of his career by Tacitus and Dio Cassius would lead one to believe. Late in life he withdrew from politics, transferred his large fortune to the imperial treasury, and devoted himself to philosophical study and writing. His suicide, after being implicated in the Pisonian conspiracy of 65, followed the model of enforced self-execution typical under the Roman emperors.

The major prose works treat a range of topics in ethics, including the theory of value and the human good, character and moral development, moral psychology, self-care and the management of emotion, friendship and political engagement, and practical morality. Occasional forays into logic and metaphysics (primarily in the *Epistles*) and physics (in the *Natural Questions*) are treated as subordinate to ethics. The intention of the mature works to represent the positions of the Stoic founders Zeno of Citium and Chrysippus is not seriously in doubt; there is little to support the claims of an older generation of scholarship for a deliberate program of eclecticism.

While Seneca claims intellectual independence, such claims are themselves in the Stoic spirit of intellectual inquiry. Doctrinal allegiance does not prevent him from studying the writings of Platonists and Epicureans, and he will occasionally endorse a point of doctrine. Epicurus himself he admires for his disciplined personal habits and effective use of various instructional methods. But it would not be accurate to describe him as amalgamating Stoic with Epicurean philosophy, for his criticism of Epicurus's hedonist foundations is as sharp as Cicero's. He is eager to claim common ground among philosophers on such points as do not require him to reverse his Stoic commitments; where definite doctrinal commitments are in evidence, however, his sympathies are decidedly Stoic.

Seneca's major work, the *Moral Epistles to Lucilius* (written between 62 and 65 CE), makes creative use of the epistolary format to present a sustained course of philosophical instruction. Intended for a wide circle of readers rather than for the sole addressee, the collection mingles scenes of daily life with a variety of topics in ethics, psychology, and occasionally metaphysics. The arrangement of ideas is deliberately unsystematic, with some topics treated in cursory fashion and others developed to considerable length. Discussions of particular interest include the non-utilitarian basis of friendship, Platonic ontology, the responsibilities of philosophers to the state, the status of moral rules, and the initial orientation of the human. A more sustained theme is the moral development of the individual as illustrated by that of the author himself and his addressee.

The earlier of Seneca's two full-length treatises, *On Benefits* (written between 56 and 64), is a study of social transactions based on similar works by the Stoics Chrysippus and Hecato. The giving and receiving of various benefactions is analyzed with rigorous attention to the motivation of the giver; there is considerable casuistic elaboration. Very different in character is the *Natural*

*Questions* (written between 62 and 65), which offers rational explanations for a list of phenomena regularly grouped in antiquity under the heading of meteorology—weather events, comets, earthquakes, and other events whose causes were not directly observable. Like Epicurus, Seneca treats such phenomena as admitting of multiple explanations; this enables him to incorporate a wide range of competing theories into his work. On the whole, however, he maintains the Stoic position on cosmic design, which he sees as having ethical significance: by pondering the regularity of the heavens and the causes of natural events, one can rise above one's ordinary objects of concern and adjust one's thought to the standard of universal reason.

Preeminent among the essays is *On Anger* (complete by 52), in the tradition of Hellenistic anger-management treatises. A careful treatment of the psychology of anger adheres to the Stoic theory of emotions generally: anger is dependent on the rational being's capacity for assent; it is intractable once begun, but can be forestalled by the techniques of cognitive therapy. Among the shorter essays, three are consolatory works of a conventional nature; the remainder treat single topics: the superiority of the virtuous person to suffering (*On Providence*) and to injury and insult (*On Constancy*); the moral end and the supposed hypocrisy of philosophers (*On the Happy Life*); remedies for spiritual malaise (*On Tranquility*); the productive use of time (*On the Brevity of Life*); and the justification for scholarly retreat (*On Leisure*).

The level of doctrinal commitment varies considerably in the shorter essays: Some restrict themselves to a Stoic viewpoint, whereas others, notably *On the Brevity of Life*, are in the spirit of generalized philosophical pro-treptic. Although the title *Dialogi* is given to the essays in the major manuscript, none is a dialogue in the sense that Plato's works are dialogues. A second speaker is sometimes made to voice an objection, but neither that voice nor the named addressee is developed into a genuine interlocutor. The essay *On Tranquility* does, however, represent its addressee as offering a confessional description of his own moral struggles. This unusual device presages the more extensive experimentation with literary form in the *Moral Epistles*.

The essay *On Clemency* has the greatest political significance of any of Seneca's works. Circulated early in Nero's reign, it celebrates what had become a watchword of the new regime and offers the essentials of a theory of good government based on the character of the ruler. Clemency, or the justifiable mitigation of justifiable penalties, is distinguished both from leniency, which is



unjustifiable mitigation, and pity, which is emotional distress at another's misfortune. Seneca attempts, characteristically, to reform Nero's administration from within, preferring failure to immediate martyrdom; in this, he differs from the hard-line idealism of Thræsea Paetus and other philosophically-minded contemporaries.

Seneca's early rhetorical training manifests itself in his style of writing, which is enhanced with clever turns of phrase, metaphor, and wit. The narrative skills displayed in his eight tragedies and in his political farce, *The Pumpkinification of Claudius*, are put to use in the philosophical prose to give effectiveness to historical anecdotes and, especially in the *Moral Epistles*, to case studies from Seneca's own acquaintance. His interest in combining philosophical with literary achievement is evidenced in his decision to write in Latin rather than Greek, which was the usual language of philosophical discourse among educated Romans of the period. It has been argued by Nussbaum and others that his tragedies are themselves experiments in ethical suasion by literary means; there is some question whether, in that case, the values promoted could remain consistent with Stoic ethics, but certainly it is true that such elements of Stoic cosmology as the cyclical conflagration do appear with some regularity in those works.

**See also** Chrysippus; Epicurus; Eudaimonia; Hellenistic Thought; Kalon; Logos; Platonism and the Platonic Tradition; Stoicism; Zeno of Citium.

### Bibliography

- Barnes, Jonathan. *Logic and the Imperial Stoa*. Leiden; New York: Brill, 1997.
- Griffin, Miriam. *Seneca: A Philosopher in Politics*. 2nd ed. Oxford: Clarendon Press, 1991.
- Hine, Harry, ed. *Naturales quaestiones*. Stuttgart: Teubner, 1996.
- Inwood, Brad. "Natural Law in Seneca." *Studia Philonica* 15 (2003): 81–99.
- Inwood, Brad. "Seneca and Psychological Dualism." In *Passions and Perceptions*, edited by J. Brunschwig and M. Nussbaum. Cambridge, MA: Cambridge University Press, 1993.
- Inwood, Brad. "The Will in Seneca the Younger." *Classical Philology* 95 (2000): 44–60.
- Nussbaum, M.C. *The Therapy of Desire: Theory and Practice in Hellenistic Ethics*. Princeton, NJ: Princeton University Press, 1994.
- Reynolds, L.D., ed. *Dialogi*. New York: Oxford University Press, 1977.
- Reynolds, L.D., ed. *Epistulae morales*. Oxford: Clarendon Press, 1965.
- Rosenmeyer, T.G. *Senecan Drama and Stoic Cosmology*. Berkeley, CA: University of California Press, 1989.

**Margaret Graver (2005)**

## SENSA

A distinction is often drawn in philosophy between two types of objects of awareness in perception. First, there are physical objects or substances (such as chairs, books, rocks, and water) and living organisms (animals, plants, and human beings insofar as they are perceptible, that is, their bodies). A common technical term for all these is *material objects*. Second, there are data of immediate awareness, which we shall refer to as "sensa" (singular, sensum), such as color patches or shapes, sounds, smells, and tactile feelings. This distinction is usually fourfold: (a) in status—material objects are external, located in physical space, and "public" (observable by different persons at once), while sensa are private and are usually held to have no external physical existence; (b) in extent—material objects may at one time correspond to several sensa and normally persist throughout the occurrence of many sensa; (c) in directness—the perception of material objects is indirect, that is, it involves inference from or interpretation of sensa that are "given" directly to consciousness; (d) in certainty—one is always certainly aware of sensa but not necessarily so of material objects.

There is no universally accepted term for sensa; *sensations* and *sense data* are commonest but indicate a further subdivision. *Sensation* is customarily used by scientists and psychologists and carries with it the suggestion that sensa are the immediate mental effects of brain activity resulting from the excitation of a sense organ by external stimuli. It and the less specialized term *sense impression* may be used interchangeably for the whole experience of awareness of sound, color, and the like, or for any sensum (such as a sound or a color patch) distinguished within it. The term *sense datum* (plural, *sense data*) apparently originated with G. E. Moore but was introduced in print by Bertrand Russell in 1912. It later became particularly associated with the sense-datum theory of Moore, C. D. Broad, and H. H. Price, while Russell developed different views and came to use other terms.

Sense data are not meant to carry any implications of causal theory, and awareness of them is called sensing (the term *sense datum* is used for the sensum only, not for the whole experience). With the development of the sense-datum theory, controversy arose between those who regarded sense data as objects distinguishable from the act of awareness of them (act/object analysis) and those who denied this and claimed that sensing is really of "sense contents" (adverbial analysis). But the terminology is generally fluid—for instance, some modern neurolo-

gists use the term *sense data* instead of *sensations* in causal contexts. Similar concepts are found in earlier writers, though their language is different. John Locke's "ideas of sense," George Berkeley's "ideas" or "sensible qualities," and David Hume's "impressions" are all forms of *sensa*.

## SENSATIONS

It has often been maintained, by philosophers as well as by psychologists, that perceiving consists in the synthesis and interpretation of sensations. But it must be realized that the occurrence of sensations in all perception is only a hypothesis and not an obvious feature of experience. In ordinary language, one may speak of having or feeling sensations of thirst, cold, or pressure and may refer to itches or pains as sensations. But the technical use of the word *sensations* involves a considerable extension of meaning, since one then speaks of visual or auditory sensations (that is, colors or sounds), while such locutions have no place in ordinary speech. We do not have green sensations in our eyes, nor do we normally feel or have sounds in our ears. Admittedly we do have afterimages, spots before the eyes, or ringing in the ears; but these are special cases because, unlike the objects or data of normal perception, the images, spots, or ringing "follows us around" and cannot be avoided by moving the head, closing the eyes, or stopping the ears. Indeed, in normal perception we are conscious not of colored shapes or of sounds as such but of material objects, or at least of ostensible material objects. Admittedly we may sometimes be aware of sounds, smells, tastes, or feelings of pressure, as distinct from objects or object properties, but it is doubtful how far these can be said to be sensations.

Sounds and smells seem public and external: Two or more people may hear the same sound or smell the same smell and agree on its source; sounds travel, and a smell may fill a room. Tastes are a borderline case—private and in the mouth, yet in a sense external to the skin and membranes—while feelings of pressure or warmth are partly sensations proper and partly seem to be awarenesses of heavy or warm objects. However, colors and colored shapes normally seem quite external, public, and at a distance from us.

Sensations in this technical sense (private mental objects of immediate awareness) are thus mainly hypothetical occurrences. Their postulation can be justified only by its success in explaining the facts of perception, and it rests on two grounds. First, there is the causal argument—perception of objects depends on and is conditioned by a chain of causal processes; for example, light waves or sound waves stimulate the appropriate sense

organ, causing impulses to travel along nerves to the brain and activate the appropriate receiving area. Perception cannot, therefore, be direct contact or confrontation with external objects—all immediate awareness must result from the causal process and be an awareness of mental sensations due to brain activity. Since they are thus separated from the external object in time and space, sensations cannot be identified with its properties, though they may resemble them.

Second, there is the psychological argument—many characteristics of perception show that it is not a direct intuitive awareness but involves interpretation of sensations. Thus, error and illusion are really misinterpretations; perception of motion, depth, and distance involves the use of sensory "cues"; and perceptual identification and discrimination are interpretative, not immediate, since they can be improved by learning and experience. (Both these arguments are discussed at greater length under the Perception entry. Here we may simply note some relevant difficulties.)

**THE EXTENT OF SENSATIONS.** Even if the causal argument forces us to distinguish between external material objects and the immediate objects of awareness caused by brain activity, it does not follow that the latter must be sensations, such as colors or sounds. They may be percepts, that is, mental contents that correspond to whole material objects, though here the psychological argument comes in, suggesting that percepts are the products of interpretation. Supporters of the theory of sensations, no doubt influenced by discoveries concerning the atomic structure of matter, at one time even claimed that the basic sensations are "atomic," that they are sensory point-elements, each corresponding to a different nerve cell—a patch of red color would thus be made up of many sensations of red. This view has now been completely abandoned, largely as the result of the experiments of the Gestalt psychologists, which show that our primary awareness is of organized wholes or figures (*Gestalten* in German), and not of elements into which these wholes might theoretically be analyzed. But even though sensations are not now thought of as minute elements that we synthesize, nonatomic sensations (colored patches of a larger size, or patterns of them, as well as sounds, smells, and so on) may still be regarded as data that we interpret in perception.

**AWARENESS AND INTERPRETATION OF SENSATIONS.** The awareness of sensations or, for that matter, of percepts must itself be explained; the danger is that it will be construed as analogous to perceiving; for example,

that seeing objects will be explained as seeing sensations caused by them, which is a circular explanation and can thus lead to an infinite regress—seeing sensations must require seeing further sensations, and so on. (Compare the duplication objection to representative realism in the Realism entry). It is therefore necessary to maintain that the awareness of sensations or percepts (“having sensations”) is a special kind of direct awareness different from perceiving, an amendment explicitly adopted by the sense-datum theory.

The problems of the psychological argument are (a) that interpretation of anything would commonly be regarded as presupposing consciousness of what is interpreted, and we are normally conscious neither of having sensations (as opposed to perceiving objects) nor of interpreting them; and (b) that the nature of the interpretation of sensations is controversial—a range of theories is possible because it is not introspectable. The sensationalists (James Mill, J. S. Mill, and others who derived their inspiration from Hume) claimed that perceiving is the association of various sensations. *Association* is a vague term and was explained as the customary linking of ideas or sensations that are similar, contiguous in space and time, and so on. F. H. Bradley and other idealists successfully attacked the sensationalist view as inadequate to explain the facts of perception; instead, they claimed that the interpretation is an inference leading to a judgment, supposing that the possibility of error in perception required this. But this overintellectualized perceiving; inferences and judgments are not the only forms of mental activity liable to error.

#### ARGUMENTS FOR THE INTRODUCTION OF SENSE DATA

Since the start of the twentieth century, philosophers have made little use of the concept of sensation in their theories but have instead talked of sense data or sense contents. Though the same things—color patches, sounds, smells, and tastes—have been put forward as examples both of sensations and of sense data, the new terminology marks several changes. Recognition of the visual depth or stereoscopic qualities of sense data means that one visual sense datum or color patch is usually held to correspond to the whole of the visible part of an ostensible object (so that one may have striped or variegated sense data). Little detailed attention has been paid to psychological phenomena, except for discussion along traditional lines of error and illusion and their bearing on whether perceiving is a form of judgment. There has also been almost a revulsion from causal arguments, clearly influenced by

their tendency to involve one in the notorious difficulties of representative realism. Instead, a fresh start has been made in the conviction that philosophy has its own distinct contribution to make in the logical and introspective analyses of perception and in the consideration of relevant epistemological issues, that is, of the extent to which perception provides knowledge of external reality. Nevertheless, with some adjustment the new arguments might be supplemented by and in turn supplement the causal and psychological arguments for sensations.

Sense data are defined as whatever is “given” or “directly present” in perceiving; they are the object of *sensing*, of “direct” or “immediate” or “actual” awareness in perception. The claim that this awareness occurs within perceiving is essential to the analysis. To most of its exponents it seems a clear fact of our experience as percipients, one revealed by reflective examination. “Direct” is explained by Price (in *Perception*) as meaning intuitive or “not reached by inference, nor by any other intellectual process.” This formal definition was often supplemented by a kind of ostensive one: Moore, J. R. Smythies, and others gave instructions for looking at an object or scene and picking out the sense datum, such as a colored shape. (Misleadingly, afterimages were sometimes offered as examples of sense data, but their difference from normal perception has already been noted; misleadingly also, some talked of seeing or hearing sense data.)

This definition of sense data naturally raises the question “Why not say that tables, chairs, and other material objects are given or directly seen?” In answering this, these philosophers produce various arguments for distinguishing sense data from material objects.

**THE CERTAINTY ARGUMENT.** The certainty argument was stressed by Price and by Russell in his search for “hard data,” though it is also found in other sources. Directness or givenness implies certainty—what is given must be limited to what we are absolutely certain of. But in any perceptual situation we cannot be sure that we are aware of any particular material object. For example, an object that seems to be a tomato may in fact be something quite different—a wax imitation, perhaps, or a reflected patch of light, or a hallucination (that is, not be a material object at all). Yet whatever the illusion may be, there can be no doubt, when we seem to see a tomato, that there is given a red, round, bulgy patch of color, a sense datum. Another version of this argument is the method of reduced claims; by confronting him with possible sources of error, you force the person concerned to reduce his

claim from “I see a tomato” to what he actually and directly sees or, rather, senses: “I see a red, round color patch.”

**THE PARTITIVE ARGUMENT.** When we observe a tomato or a bell, what we “actually see”—the “objective constituent” of the situation, what is given or sensed—is the colored shape that seems to be its front surface. This is a sense datum. We assume that the object has other surfaces and has other characteristics, such as causal properties, three-dimensionality, and persistence in time; and if we loosely say that we see a bell, we imply that we are perceiving an object possessing these properties, although we do not directly see or sense them. This argument, which stresses extent of sense experience rather than certainty, was preferred by Broad and Moore but seems inferior in suggesting that sense data are those parts of an object that we “actually see” on a given occasion—which raises difficulties with respect to illusions.

**THE ARGUMENT FROM THE CONTENT OF ILLUSIONS.** When a drunkard sees a hallucinatory pink elephant or sees two bottles when only one is present, what is the elephant or second bottle if it is nothing material? The sense-datum theorist answers, “A private object of awareness, a sense datum,” and applies this also to cases of the relativity of perceiving: For example, when a round plate looks elliptical to a person standing at one side, the elliptical appearance cannot be the plate, which is round; it is an elliptical sense datum private to that person. Indeed, it is argued that at all times we are directly aware only of sense data, since there is no qualitative jump between the cases where one cannot be directly aware of an object, and so must be sensing sense data, and the normal cases where we think we are directly aware of an object. This gradation or lack of jump is particularly clear in the case of relativity, as when we gradually move from where the plate looks round to where it looks elliptical, but it also applies to many hallucinations where the illusory *sensa* are integrated with a genuine background. In short, perceiving a material object involves sensing sense data related or “belonging” to it; when the plate looks round to me and elliptical to you, I am sensing a round sense datum belonging to it and you are sensing an elliptical one.

### THE FULL SENSE-DATUM THEORY

The fundamental conception of sense data, as directly given elements of experience, spread far beyond epistemology. Both the atomic facts of the logical atomists and the supposedly incorrigible basic or protocol proposi-

tions of the logical positivists had as their prime examples simple statements about sense data (or *sensa* generally), such as “This is red.” But the conception was also developed into a full theory of perception by consideration of the following topics, even though disagreements led to variant accounts.

**THE GENERAL NATURE OF SENSE DATA.** The arguments for the introduction of sense data, if valid, show that sense data are given and provide examples of them. Further alleged properties emerge from the discussion of illusions and relativity, namely, that sense data (1) are private, each sensed by only one percipient (see argument from the content of illusions); (2) are transitory existents, lasting only while they are sensed, so that they are usually claimed to be events rather than things or properties; (3) are distinct from the percipient and seem to be external (in contrast with sensations); (4) are without causal properties, for sounds (as opposed to sound waves) cannot act on other things, nor can colors or tastes, though the sensing of them may affect a person; (5) cannot be other than they appear to be, or the certainty argument is undermined.

Despite wide agreement on most of these points, a considerable divergence of view arose about (3) and (5). Point (3)—that sense data are distinct from the percipient and seem to be external—involves what came to be called the act-object analysis of sensing. Largely on phenomenological grounds—on how direct experience of color patches, sounds, and such seem to the person concerned—Price and others claim that sense data have distinct existence, that they are objects distinguishable from the act of awareness of them. But some philosophers maintain that the data are only “sense contents” and do not exist apart from the sensing of them any more than does a pain or sensation. This view is formulated in the so-called adverbial analysis of sensing, namely, that “I sense a red color patch” is properly to be regarded as a statement of how I sense or, to put it in a different way, “red color patch” is an internal accusative of the verb *sense*, just as “waltz” is an internal accusative of *dance* in “I danced a waltz.”

There is agreement on point (5)—that sense data cannot appear to be what they are not, for example, sense data cannot appear elliptical when they are round. (Even this is dubious—an apparently pink expanse may, on examination, be found to consist of red dots on a white background.) But some say that sense data can fail to appear as they are (do not reveal their full properties at first sight); thus, one may see that a colored datum is

striped without noting how many or how thick the stripes are. Others deny this, claiming that a closer look results in a fresh sense datum. In fact, the theory cannot deal satisfactorily with the phenomenon of attention. A thing may look quite different on careful examination from the way it looks at a casual glance, and the difference seems to be a matter of how attentively we look, a matter of changes in our mode of observation. In line with this evidence, one should say that sense data may reveal their full properties only on a closer examination, but then one is suggesting that sensing may at times be casual and inattentive and is thus undermining the fundamental claim that sensing is certain and incorrigible.

**THE RELATION OF SENSING TO PERCEIVING.** The distinction between sensing and perceiving is threefold. First, perceiving is the awareness of some material object; except in certain kinds of illusion this awareness is the result of the object in question (or light or sound from it) acting on the percipient's sense organs. Sensing is the awareness of private sense data that differ from material objects and do not affect the sense organs. Second, sensing is claimed to be direct, immediate, and incorrigible, a form of knowing. Owing to illusions, perceiving cannot be this; it is fallible and indirect. Third, the indirectness of perceiving is said to consist in its being mediated by sensing; perceiving involves sensing, contains sensing within it.

Various views are possible about the nature of this mediation of perceiving by sensing, but they are best expressed as theories of perceptual consciousness. The same kind of consciousness of a tomato, for example, seems present in normal perception, when one sees a tomato as a tomato; in an illusion, when what one sees as a tomato is a piece of wax; and in a hallucination, when no corresponding material thing is there. The kind of consciousness present in these three cases may be called perceptual consciousness and is more conveniently discussed than perceiving, where the implication that there is an object acting on the sense organs complicates the issue.

Some, such as Brand Blanshard (*The Nature of Thought*, London, 1939, Ch. 2), claim that perceptual consciousness consists in sensing a datum and judging or inferring that it belongs to a material object. Price, however, argued that this is too intellectual and does not fit the facts. We unquestioningly accept or take for granted rather than infer or judge, and therefore he defined perceptual consciousness as sensing a sense datum (or data) and taking for granted that it (or they) belong to a mate-

rial object. Others have said that we refer the sense datum to a material object, but *refer* is vague.

Two points of interest arise here. First, philosophers have most often said that we accept or judge that the sense datum belongs to a physical object. This seems obvious only about smells or tastes, and one would on first thought say we assume that the visual sense datum or color patch is the tomato. There is a reluctance on the part of sense-datum theorists to allow this, presumably because they are influenced by the partitive argument or by their knowledge that *ex hypothesi* the sense datum cannot possibly be the physical object. But there seems to be no reason why the ordinary person, whose mental processes are being described, may not mistakenly assume this; one would, for example, say "That patch of white over there on the hill is a sheep" (admittedly, the patch as "public" is hardly a sense datum, but it is the nearest one can get to a sense datum by ordinary examples).

Second, to say that we judge or infer that a sense datum belongs to (or is) a physical object is implausible, for it implies we are conscious of it first as a datum, which is not true to the facts: There is no passage of mind from datum to object, as in inference. Even to say we subconsciously judge or infer is unsatisfactory, for it seems extravagant to suppose that we constantly do subconsciously what we never do consciously. Price attempts to overcome this by maintaining that to take for granted that *A* is or belongs to *B*, one does not need to distinguish them at the time—indeed, the contrary is implied. Sensing thus comes to be regarded as a sort of sensory core within perceptual consciousness, surrounded, as it were, by the further activity of taking for granted. The two states of mind, sensing the red sense datum and consciousness of the tomato, arise together and simultaneously and can be distinguished only by subsequent analysis.

Even this account may be criticized on the grounds that it still does not do justice to the evidence of experience, namely, that perceptual consciousness is one unitary and unanalyzable state of mind, not two. No subsequent analysis of experience reveals sensing as an element within perceptual consciousness. Analysis or reflective examination can result in a "reduced" or critical phenomenological mode of observation in which one distinguishes sounds or colored shapes as such without attributing them to objects, but if this is sensing—and it seems to be the nearest one can get to it—then it is a quite different state of mind from normal perceiving. There is no ground for supposing that this, achievable only by an

effort of analysis, occurs as part of normal unconsidered perception. In general, therefore, the attempt to establish sensing sense data as an omnipresent basic element in perceiving faces the same difficulties that faced the claim that perceiving is the interpretation of sensations.

Another way of seeing the error is to consider the normal usage of “taking for granted.” Price’s analysis is at first sight closest to “*Y* saw the book and took for granted that it belonged to *B*,” but then *Y* is referred to as conscious of the book, while the average percipient is not conscious of sense data as sense data; he is conscious only of the material object. This difficulty can be avoided by the formulation “*X* took for granted that *A* was *B*”; for example, that the piece of wax was a tomato, or that the visitor was the man he was expecting. In each case both *A* and *B* denote the same entity (the wax or the visitor). *A* describes this entity in a way that the speaker knows to be correct; *B* describes it as *X* saw it. Similarly, one might say, “He took for granted that the sense datum (*A*) was a material object (*B*).” But this will not really save the analysis in which the datum and the physical object are alleged to be two quite different entities; to fit the analysis the first phrase (*A*) must also be a description of the alleged object of awareness of *X*, not of the speaker. Price seems to be making the mistake of offering as a description of a percipient’s actual mental content what is in fact a description of the situation that can be made only by someone correcting the percipient’s error.

**THE RELATION OF SENSE DATA TO PHYSICAL OBJECTS.** One of the vaunted advantages of the sense-datum analysis of perception is its neutrality with respect to the traditional realist theories of knowledge. (Idealism was ruled out by the original claim that sense data are distinct from the sensing of them.) Indeed, sense data were even said to be neutral in that so far as the analysis is concerned, they can be mental or physical or neither. Consequently, it is possible to state the various theories of knowledge in terms of sense data. Naive realism reduces to the view that sense data are parts of the surface of material objects; representative realism would claim that sense data are mental existents caused or generated by cerebral activity ultimately due to material objects and that sense data resemble the properties of these objects. (The second view and, if not too naive, the first also, could admit “wild sense data”—hallucinations that are not part of or caused by physical objects.)

Moore at times toyed with supposing that sense data are parts of the surface of objects (and even seriously discussed whether they might be identical with objects),

though this must have been due to his affection for the partitive argument. The other arguments for sense data and general considerations about illusion do not allow this; for example, a round dish cannot have an elliptical sense datum as part of its surface. Representative realism is a more likely possibility: Neurologists such as Smythies advocate this theory in terms of sense data, and Broad proposed something not unlike it. Most of the philosophers have, however, rejected it in view of its traditional difficulty—if our observation is limited to sense data while material objects are only assumed causes of sense data, then these objects are in fact never observed and therefore may, for all we know, not really exist.

A more common view is that sense data belong to material objects in the special sense that the latter are composed of “families” of sense data. This “family” relationship is not literally one of whole and part, as in naive realism; the material object is supposed to be a complex system or pattern of groups or sequences of sense data. But if a physical object is simply a family of sense data, then when no sense datum occurs—when the object is unobserved—the object must cease to exist. This is felt to be too paradoxical, and two main lines of development within this view have been put forward: (1) phenomenalism, in which the object is regarded as a family of actual and possible sense data—when unobserved, it consists solely of possible sense data; (2) a compromise theory put forward by Price in which the material object, while mainly such a family, contains a physical occupant that persists, even while it is unobserved, as the source of all its causal properties. The notion of a physical occupant has some analogies to Immanuel Kant’s notorious thing-in-itself, and this view has not obtained widespread acceptance.

This divergence of view reflects a central dilemma in the sense-datum theory. If the theory maintains that sense data belong to material objects or that the latter in some way consist of them, then it is difficult to explain (*a*) the persistence of such objects when unobserved; (*b*) the privacy that all versions attribute to sense data—how can a public object be a family of private sense data?; (*c*) the conditioning or even generation of sense data by the sense organs and nervous system, which is required by the physiological facts, by the occurrence of hallucinations or color blindness, and by the effects of attention and learning on perception. (Most sense-datum theorists admit the generation as well as the conditioning.) But if one does not say that sense data belong to or constitute material objects, the distinctness and apparent depth of sense data (at least of visual ones) is difficult to explain;

and, more important, sense data tend to become mental entities like sensations. This, together with the privacy and the generation by the brain, leads one into representative realism.

One attempt to escape this dilemma is to say that sense data are extended and located in their own private “sensible” space along the lines first suggested by Russell in his *Mysticism and Logic* and *Our Knowledge of the External World*. There is one such sensible space, with its own extension and dimensions, for every point in physical space, and the latter in fact becomes the system of points at which sensible space occurs. A physical object is thus, as it were, spread over physical space in a series of “perspectives” or “unperceived aspects,” in the special sense that from different points in physical space, granted that sense organs and brain function properly, sense data may occur in sensible space but also belong to the object as appearances of it and reproduce its characteristics in a way modified both by the viewpoint and by the nature of the sensory apparatus.

This theory is very complex, which means that any summary of it is necessarily garbled. Two of the complexities are that a special interpretation is needed of what we normally call the volume occupied by a physical object and that account must be taken of the different senses, for sight, sound, and touch at least each have their own specific spaces. (Russell later spoke of sensible space as a construct of these spaces, but a construct cannot be the space in which immediately given sense data are located.) A further difficulty is that a given sensible space cannot really be at a point. Not only are the hands, say, at some distance from the eyes, but the brain and the sensory activity associated with perception of an object at one time and place are also really spread over an area. However, the major objection is once again the causation and conditioning of sense data by sense organs and nervous system. How do they influence or produce data in sensible space, or modify the appearance in sensible space of an object in physical space? As soon as one tries to fit in the causal processes, it is difficult to avoid straightforward representative realism, in which all this elaboration becomes unnecessary; perspectives become otiose, except as mere possibilities, or turn into light waves and sound waves. Hence, Russell’s later views gradually approach representative realism (for example, in *Human Knowledge*, 1948).

There does, in fact, seem to be no satisfactory way out of this dilemma for the sense-datum theory. Upholders of it must embrace one horn or the other—they must maintain pure phenomenalism or representative realism. Each has its well-known difficulties, but the second,

though once thought hopeless, is now perhaps more easily made plausible than the first.

## DIFFICULTIES CONCERNING SENSE DATA

A number of difficulties have been noted already in the full theory, but others lie even in the arguments for sense data.

**THE CERTAINTY ARGUMENT.** Various objections may be made to the certainty argument. First, so far as introspective examination is concerned, our awareness is, as we have mentioned, of putative objects, not of color patches—one sees a tomato or something looking like one. Awareness of color patches as such is a different kind of observation from normal perceiving, not a sensory core within it. One may more readily be said to be directly aware of sounds or smells as such; but even then, as we saw concerning sensations, one is aware of them as public and external, not private.

Second, the assumed link between immediacy and certainty is questionable. If immediacy is put forward as an introspective characteristic of the awareness of sense data, nothing follows about its certainty because any awareness we point to as direct may be mistaken. However, if immediacy and certainty are linked conceptually, as the premise of the certainty argument suggests—if they are defined in terms of each other—then it may be that what seems to be immediate, and hence certain, awareness is not immediate. This point may be illustrated in various ways. The certainty argument claims that sensing reveals existents—that when we look at an (apparent) tomato, we cannot doubt that something red and round and bulgy exists. Strictly speaking, however, we are certain only of something red-looking; it may in fact be orange that looks red in this light. Indeed, as J. L. Austin pointed out, even statements about how a thing looks may have to be retracted. Further, the controversy over whether sense data can fail to appear as what they are throws further doubts on the incorrigibility claim, and the alternative adverbial analysis, that sense data are only sense contents, challenges the view that something exists distinct from the percipient’s experience of it.

Third, the certainty argument is too ready to deny that we see physical objects in cases of illusion and distortion and to assume that we are aware of the same kind of existent in both perception and hallucination. Both these assumptions may plausibly be denied. When we look at the putative tomato, even if it is a piece of wax or a reflection of a tomato or an image on a screen, we are

still seeing a material object—wax, or the tomato “in” (via) the mirror, or a screen illuminated in a certain way. There is no need to suppose that we are aware of something else, a sense datum. Contrastingly, the common explanation of hallucinations would be that they are unusually vivid mental images confused with perceptions. Such images, like afterimages, seem to be private, but one should not assume that they are identical with what we are aware of in normal perception. The sense-datum theory can, however, reply that hallucinations are normally quite indistinguishable by the victim from normal perception and may also be integrated with a perceived background—for instance, the apparition may walk across the room and cast shadows—so if the hallucinatory images are private, so must be the data of the background. Although two entities are not necessarily identical because they are generally indistinguishable, identity may be the most plausible explanation of their indistinguishability, and the integration is very difficult to explain except on the sense-datum theory or on some form of representative realism. All the same, the sense-datum theory, if treated as an explanatory hypothesis, has the disadvantage of being very uneconomical in postulating so many distinct entities (the sense data).

**THE PARTITIVE ARGUMENT.** The partitive argument can be dismissed quite briefly, apart from its other troubles already mentioned. From the fact that we do not actually see the whole of an object at once, it does not follow that we do not then see the object, any more than the fact that we cannot visit all of New York at once means that we cannot visit it at all. Consequently, there is no ground for regarding what we actually see of an object as something different from it (a sense datum) or the actual seeing as some special direct awareness (sensing).

**THE ARGUMENT FROM THE CONTENT OF ILLUSIONS.** The argument from the content of illusions presents problems similar to those of the certainty argument. The alternative to the sense-datum answer concerning what the drunkard sees in hallucinations is “a mental image,” and in double vision “one bottle looking double.” Neither answer is wholly satisfactory, since the first cannot explain the integration of the image with a real background, and the second has been accused of evading the issue—looking double is not like looking blue or looking elliptical, for it involves an extra apparent object, not a differing quality of the one object.

Ordinary cases of relativity are much more easily dealt with. When one sees a round dish that looks elliptical, one is simply seeing the dish and not some elliptical

existent; the theory oddly assumes that things cannot look other than they are. This assumption is linked with the notion of immediacy: It is gratuitously supposed that in seeing the dish as elliptical, one is immediately aware of an elliptical existent. However, this begs the question by equating immediacy with incorrigibility, so that what looks elliptical is said to be elliptical. Furthermore, there is no cogent ground in experience or in the argument for supposing that nonhallucinatory sense data are private to a person: The elliptical shape of the plate or even the second bottle might also be sensed by others. The privacy is best supported by arguing that sense data are “generated” by brain processes (as in the causal argument for sensations).

**OTHER DIFFICULTIES.** Various other criticisms of sense data have been put forward, especially by Gilbert Ryle and J. L. Austin. First, sensing is either seeing under another name—in which case there is the reduplication or regress noted concerning sensations—or else it is a myth. The notion of a mistake-proof awareness, Ryle claimed, arises from misunderstanding the character of perception words, which are achievement words or indicate the scoring of an investigational success. One cannot perceive unsuccessfully any more than one can win unsuccessfully, but that is a linguistic or conceptual matter; it does not mean that if one looks or plays, one is bound to see or to win.

Second, the theory, in speaking of sense data as existents, is simply reifying (treating as things) the sounds, smells, or looks of things. Ryle claimed a linguistic origin for this: By wrongly speaking of “seeing looks” or “smelling whiffs,” which are pleonastic usages like “eating nibbles,” the theory tends to treat looks and whiffs as the sort of things we can see or smell—that is, as objects—and fails to see that the point of such words is to show how we are perceiving objects. (He could hardly condemn hearing sounds, even if the other examples are correct.)

Third, Austin attacked the tendency of Moore, A. J. Ayer, and others to distinguish different senses of the word *see*: the normal sense (seeing objects) and the restricted “direct” or “actual” seeing (sensing, which is incorrigible). He claimed that the second sense is a myth: The basic fact is that one may describe the object one sees in various ways, depending on how advertent one is; for example, as a tomato or as a red object. But in these two cases it is the same thing described in different ways, not two different things; nor does it follow that there are two kinds of seeing or two senses of the word *see*.



Austin had other alleged linguistic grounds for the theory's mistakes, such as confusion of illusion with delusion, but it is doubtful whether the several different linguistic origins that he and Ryle claimed for the theory are really genuine and important. The reflective examination of experience seems a more likely origin for the theory, in view of the stress laid on it by the sense-datum philosophers. They have been so struck by the apparent immediacy of perceiving, by its apparently direct confrontation with a "given," that they have readily assumed that it does involve such an immediate awareness or confrontation; and because (on account of illusions) they cannot identify immediate awareness with the perception of a physical object, they have supposed it to be an inner awareness of special data—the sensing of sense data.

### FURTHER DEVELOPMENTS

**SENSE CONTENTS.** As we have seen, the adverbial analysis of sensing claims that *sensa* no more exist as entities distinct from the sensing of them than do itches or pains; consequently, they are often referred to in this analysis as sense contents. Important advocates of this approach have been C. J. Ducasse and Ayer, and under its influence Moore modified and Russell abandoned an earlier faith in an act—object analysis (that sense data are separate entities distinct from the act of sensing). Russell's conversion to the adverbial analysis was brought about by his conclusion that the subject of awareness is a logical fiction; since the act-object analysis presupposes a subject of the act of awareness, it had to be dismissed (*Analysis of Mind*, p. 141). Probably few would follow him on this; it is, at any rate, not clear that the adverbial analysis can dispense with the subject, nor is it clear why one should wish to. Moore's *Refutation of Idealism* relied on the act-object analysis, but he later had doubts about this. He tended to see the problem as whether sense data have any existence when unperceived, or rather, unsensed; that is, whether their *esse* is *percipi* or *sentiri*. He regarded this as an open question, producing various arguments on either side at different times. Actually, the two questions are not quite identical: The adverbial analysis implies that sense data or sense contents cannot exist unsensed, while the act-object analysis is neutral on this.

It seems clear that whether sense data exist unsensed is not a question that can be settled by sensing them. Consequently, some would say it is a purely conceptual matter, one of how sense data are to be defined or how the general theory is to be framed. But factual issues are relevant and present a dilemma similar to the one of the relation of sense data to physical objects. If one accepts

that sense data are generated by the brain, then it seems that they cannot exist unsensed. Even if they are only conditioned by the nervous system, they must appear different from what they really are in the unconditioned, unsensed state, thus undermining the certainty argument. At the same time, to say that a physical object is a family of sense data is scarcely meaningful if sense data do not exist unsensed; therefore, Russell at one period claimed that they do exist unsensed, calling them *sensibilia* in this state. More usually, however, phenomenalism is maintained; sense data do not exist unsensed, but possible ones or possibilities of sensation do.

So far as introspection is concerned, decision between the analyses depends on which sense is considered. Visual sense data, such as color shapes, would seem clearly to be distinct and to require an act-object analysis. (Afterimages are more doubtful, but anyhow are a special case.) Much the same applies to sounds and smells, which are normally experienced as external: By contrast, tactile and other bodily (somatic) sense data, such as pains or feelings of warmth or pressure, and the sensations of movement (kinesthetic data) seem clearly adverbial, as perhaps is taste; but there are marginal cases. Explanation of this variation is difficult for the theory, which would be more plausible if it could give one account of all sense data; it is also difficult to square the distinctness claimed in the act-object analysis with the privacy always claimed. Another possible line, which seems required for dreams and mental images and for hallucinations where no distinct objects are present, would be to say that while sense data seem to the person to be distinct, they are actually contents of adverbial experiences, as are sensations. However, this would undermine the claim of the theory to rely on introspective analysis.

**SENSE-DATUM LANGUAGE.** One suggestion that has been made is that the sense-datum philosophers have not, as they at first thought, produced a new theory of perception; they have simply introduced a new and more convenient terminology for discussing the facts of ordinary perceiving. This was accepted for a time by those who sought to see all philosophy as dealing with language and by those who, impressed by the difficulties the sense-datum theory encountered, sought to salvage something from the wreck. It is not popular now, for those with a linguistic bias have turned to the examination of ordinary language rather than to the advocacy of new terminologies, while the general decline of support for sense data has proceeded beyond this halfway house. Another reason for supposing that the sense-datum theory was only a terminology was the view that theories must be verifi-

able by observation of predicted consequences, which the sense-datum theory is not, but this seems to confuse a philosophical theory with a scientific theory.

Considered simply as a terminology, the language of sensing and sense data was claimed to have certain advantages; for example, that it is (a) noncommittal—one can describe the contents of one's experience independently of the physical objects they are thought to refer to—and (b) neater than ordinary language, for one can avoid periphrases like “there appears to be a red, bulgy tomato-like object” merely by listing the data sensed. But these are only slight advantages, and it seems that they are far outweighed by the fact that a sense-datum language cannot be truly neutral. It has been so long associated with the sense-datum theory that it must inevitably beg the question by suggesting that the data are private, transitory existents; that one is not “actually seeing” physical objects; or that in describing the scene in terms of visual and tactile data, one has described the experiences of normal perception and not of the different “reduced” phenomenological observation.

**See also** Perception; Phenomenalism; Realism.

## Bibliography

### SENSATIONS

Readers are advised to consult the historical surveys and then follow up any particular treatment that interests them with the aid of the references given. On the psychological side, see Edwin Garrigues Boring, *Sensation and Perception in the History of Experimental Psychology* (New York: Appleton-Century, 1942) and *A History of Experimental Psychology*, 2nd ed. (New York: Appleton-Century-Crofts, 1950); on the philosophical side, most useful is David Walter Hamlyn, *Sensation and Perception* (New York: Humanities Press, 1961). A stimulating but idiosyncratic discussion is given in Bertrand Russell, *The Analysis of Mind* (London: Allen and Unwin, 1921). Also, many of the books listed below under sense data discuss sensations. Gilbert Ryle's article “Sensations,” in *Contemporary British Philosophy*, edited by Hywel David Lewis, 3rd series (London, 1956), attacks the whole concept; a careful study of a limited field is David Malet Armstrong, *Bodily Sensations* (London: Routledge and Paul, 1962).

### SENSE DATA

The following contain statements of the sense-datum theory. Pioneer work is found in George Edward Moore, *Some Main Problems of Philosophy* (introductory lectures of 1910–1911; London: Allen and Unwin, 1953) and his more difficult *Philosophical Studies* (London: Routledge, 1922). He kept returning to the problems of perception, not always consistently; for the advanced student, Alan Richard White, *G. E. Moore* (Oxford: Blackwell, 1958), is a useful guide with a full bibliography. Bertrand Russell was another pioneer in his lucid elementary *Problems of Philosophy* (London:

Williams and Norgate, 1912) and his more significant *Mysticism and Logic* (articles of 1914; London: Allen and Unwin, 1918) and *Our Knowledge of the External World* (London: Allen and Unwin, 1914), but he tended to subordinate the topic of sense data to his special perspective theory. Clear and systematic is Henry Habberley Price, *Perception* (London: Methuen, 1932), which develops a full sense-datum theory. Another well-known account is Charlie Dunbar Broad, *The Mind and Its Place in Nature* (London: Kegan Paul, 1925); earlier and fuller statements of his views, with more attention to causal problems are in his *Scientific Thought* (London: Kegan Paul, 1923) and *Perception, Physics and Reality* (Cambridge, U.K.: Cambridge University Press, 1914). Alfred Jules Ayer gives several clear discussions, mainly from a phenomenalist point of view, in *The Foundations of Empirical Knowledge* (London: Macmillan, 1940), *Philosophical Essays* (London: Macmillan, 1954), and *The Problem of Knowledge* (London: Macmillan, 1956). For a modern version of representative realism stated in terms of sense data, see John Raymond Smythies, *Analysis of Perception* (London: Routledge and Paul, 1956).

There are many criticisms of the sense-datum theory. A general introductory survey, chiefly of Price's version, and extended criticisms are given in Rodney Julian Hirst, *The Problems of Perception* (London: Allen and Unwin, 1959). Gilbert Ryle's criticisms are found in his *The Concept of Mind* (London: Hutchinson, 1949) and *Dilemmas* (Cambridge, U.K.: Cambridge University Press, 1954); John Langshaw Austin's are in his lively defense of a commonsense approach, *Sense and Sensibilia* (London: Oxford University Press, 1962); Roderick M. Chisholm has some succinct criticisms in his more technical *Perceiving: A Philosophical Study* (Ithaca, NY: Cornell University Press, 1957); Harold Arthur Prichard, *Knowledge and Perception*, edited by W. D. Ross (Oxford: Clarendon Press, 1950), attacks Russell and the view that sensing is a form of knowing. Martin Lean, *Sense Perception and Matter* (London: Routledge and Paul, 1953), attacks Broad's version; and valuable is H. H. Price's “The Nature and Status of Sense-Data in Broad's Epistemology,” in *The Philosophy of C. D. Broad*, edited by Paul Arthur Schilpp (New York: Tudor, 1959). On Moore, besides A. R. White's book, senior students should see P. A. Schilpp, ed., *The Philosophy of G. E. Moore* (Evanston, IL: Northwestern University Press, 1942), particularly C. J. Ducasse, “Moore's Refutation of Idealism”; O. K. Bouwsma, “Moore's Theory of Sense-Data”; and Paul Marhenke, “Moore's Analysis of Sense Perception”; and Moore's replies to them.

Among many critical articles, notable are Roderick Firth, “Sense-Data and the Percept Theory,” in *Mind* 58 (232) (1949): 434–465, and 59 (233) (1950): 35–56, attacking the sense-datum analysis of perceptual consciousness; Winston H. F. Barnes, “The Myth of Sense-Data,” in *PAS* 45 (1944–1945); and Anthony M. Quinton, “The Problem of Perception,” in *Mind* 64 (253) (1955): 28–51, criticizing the concept of sense data and the arguments for them; also a more advanced discussion of the earlier versions of the sense-datum theory in the symposia “The Status of Sense-Data,” in *PAS* 14 (1913–1914): 355–406, and “The Nature of Sensible Appearances,” in *PAS*, Supp. 6 (1926): 142–205.

## SENSE-DATUM LANGUAGE

The linguistic interpretation of sense data is developed in George A. Paul, "Is There a Problem about Sense-Data?," in *PAS*, Supp. 15 (1936), and A. J. Ayer, "The Terminology of Sense-Data," in *Mind* 54 (216) (1945): 289–312, reprinted in his *Philosophical Essays* (London: Macmillan, 1954); there is a useful discussion of it in the symposium "Seeming," in *PAS*, Supp. 26 (1952): 195–252.

## SENSE CONTENTS (ADVERBIAL ANALYSIS)

The most important advocate of this approach is C. J. Ducasse, *Nature, Mind and Death* (La Salle, IL: Open Court, 1951), largely identical with his contribution to *The Philosophy of G. E. Moore*. Brief statements may be found in A. J. Ayer, *Language, Truth and Logic* (London: Gollancz, 1946) and *Philosophical Essays* (London: Macmillan, 1954). A fuller discussion, with emphasis on the differences between the senses, is C. D. Broad, "Berkeley's Argument about Material Substances," in *Proceedings of the British Academy* 28 (1942): 127ff., and his *Scientific Thought* (London: Kegan Paul, 1923).

## OTHER RECOMMENDED TITLES

- Anscombe, G. E. M. "On Sensations of Position." *Analysis* 22 (1962): 55–58.
- Armstrong, David Malet. *Perception and the Physical World*. London: Routledge and Kegan Paul, 1961.
- Casullo, Albert. "A Defense of Sense-Data." *Philosophy and Phenomenological Research* 48 (1987): 45–61.
- Chisholm, Roderick. *The Foundations of Knowing*. Minneapolis: University of Minnesota Press, 1982.
- Chisholm, Roderick. "The Problem of the Speckled Hen." *Mind* 51 (1942): 368–373.
- Cornman, James, "Sellars, Scientific Realism, and Sensa." *Review of Metaphysics* 23 (1969): 417–451.
- Fales, Evan. *A Defense of the Given*. Lanham, MD: Rowman & Littlefield, 1996.
- Fantl, Jeremy, and Robert Howell. "Sensations, Swatches, and Speckled Hens." *Pacific Philosophical Quarterly* 84 (2003): 371–383.
- Huemer, Michael. "Sense-Data." Stanford Encyclopedia of Philosophy. Available from <http://www.plato.stanford.edu>.
- Jackson, Frank. *Perception: A Representative Theory*. Cambridge, U.K.: Cambridge University Press, 1977.
- Lowe, E. J. "Indirect Perception and Sense Data." *Philosophical Quarterly* 31 (1981): 330–342.
- Martin, Michael G. F. "Beyond Dispute: Sense-Data, Intentionality, and the Mind Body Problem." In *History of the Mind-Body Problem*, edited by Tim Crane. New York: Routledge, 2000.
- Perkins, M. *Sensing the World*. Indianapolis: Hackett, 1983.
- Sellars, Wilfrid. "Phenomenalism." In his *Science, Perception and Reality*. London: Routledge, 1963.
- Swartz, R. J., ed. *Perceiving, Sensing, and Knowing*. New York: Doubleday, 1965.

**R. J. Hirst (1967)**

*Bibliography updated by Benjamin Fiedor (2005)*

## SENSATIONALISM

"Sensationalism," the doctrine that all knowledge is derived from sensations, takes several closely related forms. As a psychological theory it stresses the origins of knowledge and the processes by which it is acquired; it seeks to reduce all mental contents to unitary sensations and has close connections with associationism. It is sometimes, as by its acute but sympathetic critic James Ward, called presentationism. As an epistemological theory it tends toward the view that statements purporting to describe the world are analyzable into statements concerning the relations between sensations and that this analysis elucidates the meanings of the original statements. It is sometimes regarded as a form of empiricism and adopted with antimetaphysical intentions.

Sensations are usually regarded as occurrences in us, either caused by external objects (Epicurus and John Locke) or not meaningfully attributable to external causes (James Mill and Étienne Bonnot de Condillac). By some they are explicitly likened to feelings or emotions (Anaxagoras and David Hartley), and by others to images (Ernst Mach); the more modern forms, however, probably depend, even if not explicitly, on taking them all as analogous to feelings.

There is a tendency to associate sensationalism with the nineteenth and twentieth centuries, as a development of the work of the empiricists of the seventeenth and eighteenth centuries, but it actually has a long history. A study of its development takes us back to the pre-Socratics, and although in its modern forms it usually leans heavily on the distinction between sensation and perception, there were views that can be called sensationist long before the distinction was made (for example, Protagoras held them). The distinction between sensation and perception is used because it is believed that although perception involves interpretation and, thus, the possibility of error, sensation does not. Sensationalism is therefore sometimes looked upon as the end point of the empiricists' quest for certainty and a sure foundation for knowledge.

## THE GREEKS

The Greeks had no linguistic means of distinguishing between sensation and perception, but they do not appear to have considered this a serious lack. The pre-Socratics were apparently interested in perception mainly from the physiological and physical point of view; they wanted to describe processes, which they tended to see as purely mechanical (this is especially true of Empedocles

and the atomists), involving the meeting of effluences from the sense organ and the external object. But Anaxagoras introduced a feature that has some significance for an understanding of sensationalism—namely, the idea that perception involves pain. This facilitates the assimilation of all sensations to feelings referred to below.

**PROTAGORAS.** Protagoras, accepting the Heraclitean view that all is change or becoming and having concluded that “man is the measure of all things,” found it easy to regard our constantly changing sense experiences as the objects of knowledge and to hold that all the so-called qualities of things, not merely the secondary qualities as the atomists believed, were relative to the perceiver. This turned attention to epistemological questions connected with the nature of perception.

**PLATO.** Perhaps Plato and Aristotle were primarily reacting against this view of Protagoras in their discussions of perception. Plato’s argument in the *Republic* is that sense experience does not give knowledge but only opinion, since knowledge must be certain and cannot be of what is constantly changing—that is, sensations or the sensible world. According to some scholars—D. W. Hamlyn, for example—another view can be extracted from the later dialogue the *Theaetetus*, but this is highly controversial. Protagoras was referring to knowledge of a familiar, everyday sort. The view allegedly to be found in the *Theaetetus* is that the senses can give us this rudimentary empirical knowledge; they give us direct acquaintance with the outside world and even without interpretation can therefore give us knowledge. There is no distinction to be made, as far as the sensible world itself is concerned, between what is and what appears. Because sense experience is caused by the external world, it can be regarded as infallible. But this step is suspect both on general grounds and in relation to Plato’s own insistence that the categories of right and wrong are contributed by the mind. His thought seems to be that if judgment is made by the mind and if saying that something is wrong is making a judgment, then bare sense experience, being prior to judgments of it, cannot ever be said to be wrong. It should, of course, be added that it cannot be said to be right either.

**ARISTOTLE.** Aristotle, in attempting to refute the sensationalism of Protagoras, stressed the element of judgment in perception and almost arrived at the distinction between sensation and perception. At the same time he appears to admit an important feature of sensationalism. Each sense has its proper object or special sensible; the

proper object of hearing is sound and that of sight is color. But there are also common sensibles, qualities of objects that are not specially related to any one of the five senses but that are related to the common nature of them all, which he referred to as the common sense. These qualities are, roughly, the primary qualities motion, rest, shape, size, and number. Because there is a necessary connection between each sense and its special sensible, it is impossible for the senses to make mistakes about them; for example, hearing cannot err about the fact that it is concerned with sound and not color. This, however, does not entail any incorrigibility in the deliverances of the senses as is required by sensationalism. It simply means that each sense is necessarily concerned with its special sensible. Aristotle’s claims about incorrigibility probably arise, as Hamlyn says, from an unresolved conflict between his view of the senses as both active and passive. (The senses can make mistakes only if they are active and make judgments; as mere passive receptors, they cannot. If we fail to distinguish in this way, we may think of the senses as judging infallibly.) In *De Memoria et Reminiscencia* Aristotle outlined some principles of association that look forward to later accounts.

**EPICURUS.** Epicurus, who believed that sense perception is the source of all knowledge, held a causal theory of perception. He did not distinguish between sensation and perception and regarded what were later called sensations as incorrigible because caused. He was an atomist and attempted a mechanical account of perception. The Stoics opposed this account and again stressed the importance of at least rudimentary judgment in perception. Their conception of *phantasiae* roughly corresponds to the conception of sensations as images; they held that these were not necessarily veridical although some of them were intuitively certain.

## THE SCHOLASTICS

Problems of perception were not central in medieval philosophy except as they bore on the relation between empirical and other varieties of knowledge.

**AUGUSTINE.** Augustine is important on the subject of perception perhaps only because he saw that it is not meaningful to talk of sensations as either true or false; these terms can be applied only to judgments. He simply assumed that sense impressions correspond to the external world but regarded the knowledge thus obtained as of the lowest kind.

**THOMAS AQUINAS.** Thomas Aquinas followed Aristotle in his views on perception to the extent of holding that it involves the reception of a sensible form without matter, but this produces a change in the soul, not merely, as for Aristotle, in the sense organ. Sensory images (*phantasmata*) are received passively, but they are images of external objects. They have the peculiarity that we are not aware of them. The mind abstracts universal qualities from these and uses them in making judgments. The senses and the intellect are closely connected: *Nihil est in intellectu quod non prius in sensu* (Nothing is in the intellect that was not first in the senses). Because our perceptions involve judgments, they may or may not be veridical, but the *phantasmata* are not appropriately called either. This, with the fact that the *phantasmata* are images of something, prevented Thomas from being a sensationalist, but he was very close to being one in spirit and utterance.

**OCKHAM.** Although William of Ockham differed from Thomas in many ways, he also distinguished a sensible and an intellectual element in cognition. Those cognitions that involve only immediate experiences are said to be perfect. Error arises in judgment, but when we are directly apprehending something, we are not in error.

## THE SEVENTEENTH AND EIGHTEENTH CENTURIES

Sensationalism proper can perhaps be regarded as the product of a steady development of empiricist ideas from the seventeenth century to the nineteenth. Thomas Hobbes is sometimes credited with its inception, but his sensationalism is rudimentary. He did have some conception of the association of ideas and, of course, contributed to the foundations of empiricism.

Largely because of the climate of scientific opinion, involving as it did a growing belief in the importance of observation and experiment, the philosophers of the seventeenth century were much concerned with problems of perception. They were especially interested in the elimination of errors arising from sense experience and in the attempt to make our knowledge of the natural world as reliable as possible. The rationalists attempted to show that knowledge could be based on indubitable truths of reason, independent of sense experience. The empiricists sought a hard core of indubitable truths involved in sense experience upon which all knowledge could be based.

**GALILEO.** Galileo Galilei distinguished between primary and secondary qualities and thought that secondary qual-

ities existed only as sensations in us. They are, however, caused by primary qualities in objects, especially by shape and motion.

**HOBBS.** Under the influence of Galileo, René Descartes, Marin Mersenne, and Pierre Gassendi, Hobbes developed the philosophy of motion into what must be the most thoroughgoing materialism there has ever been. For him all our inquiries must start from sense experience, but there are certain principles—for instance, that motion cannot be understood to have any other cause besides motion—which we know independently of sense experience and upon which other knowledge depends. Nothing exists but matter in motion, so sensations are material changes in us that somehow mediate between motions in the external world and the minute motions of our bodily parts. Hobbes assumed the existence of external motions causing our sensations; knowledge of these “objects” can come only through sensations. This does not entail the empiricist view that all knowledge is reducible to knowledge of sensations; Hobbes was in general a rationalist, for he held that certain truths of reason are essential even for that knowledge of the natural world which depends upon sensation.

**LOCKE.** Locke’s work marks the beginning of the growth of sensationalism proper, although he was not himself a sensationalist just because he did not develop his particular form of empiricism consistently. His “ideas of sensation” are close to what were later called simply sensations, but his representative theory of perception and his assertion of the existence of substance entail that in spite of explicit claims he relied on knowledge which did not come entirely through sensation.

**BERKELEY.** George Berkeley attempted to remove this inconsistency in his attack on material substance and representative perception. Whether we view his reliance on God as the unempirical importing of a concept merely for the purpose of filling an embarrassing gap—that is, to allow us to hold that objects continue to exist when no human being is perceiving them—or as the attempt to delineate a concept that is logically necessitated by our experience, Berkeley’s account of ideas brings us very near sensationalism. There is no talk of external objects that are composed of any material different in kind from what we directly know—that is, ideas. Later sensationalism can be regarded as comparable to Berkeley’s system without God, with all its problems as well as its advantages.

**HUME.** David Hume continued this development, in one direction by rejecting mental substance, which was retained by Berkeley, as well as material substance. The world for us, as far as we can justifiably say in philosophical contexts, consists of impressions and ideas, and knowledge is of relations between these. Hume was not, however, as great a skeptic as is often alleged. We have, naturally, certain beliefs—for example, in the external world and in causal efficacy—which cannot be rationally supported. When philosophy fails to provide this rational support, so much the worse for philosophy. If Hume had not been affected by the common view that knowledge implies certainty, he would no doubt have admitted these “natural” beliefs as knowledge and thus have been farther from sensationalism in his official theory than he actually was.

Sensationalism in its fullest sense is best seen in the works of Hume’s lesser-known contemporaries Hartley and Condillac. Hartley’s work was later developed by James Mill, and its most thoroughgoing exponent in the nineteenth century was perhaps Mach.

**HARTLEY.** Hartley was a medical man; his interests were largely physiological, and his work stimulated the development of a school of psychology. His basic concepts were sensations and the association of ideas, for which he admitted a debt to Locke and Isaac Newton. All mental occurrences originate in sensations caused by vibrations of minute particles of the brain set off by external stimulation. Simple ideas are “copies” of sensations—that is, physiologically they are tiny vibrations corresponding in character to the original vibrations and left behind by sensations when the stimulus is withdrawn. Complex ideas are built up from these by association according to certain discoverable principles. The vibrations occur in a subtle elastic fluid in the medullary substance of the nerves and brain. This mechanical account is reminiscent of Hobbes’s view and admittedly owes a debt to Newton’s mechanistic philosophy. The conception of the association of ideas springs from Locke, and the consequent contention that ideas are copies of sensations echoes Hume’s account of impressions and ideas. Hartley’s theory leads to the conclusion that we are aware only of occurrences within ourselves but that these depend for their character on the external world. There is a twofold correspondence, between ideas and sensations and between sensations and stimuli.

**JAMES MILL.** James Mill accepted Hartley’s basic conceptions and developed the psychological side of the theory. Hartley had expressed in terms of vibrations two princi-

pal determinants of the strength of association—the vividness of the sensations and the frequency of their conjoint occurrence. Mill discussed these principles in some detail, without Hartley’s preoccupation with vibrations, contrasting his principles of association with Hume’s and using some rather unsatisfactory arguments for preferring his own. In place of Hume’s contiguity in time and place, causation, and resemblance, Mill put synchronous order and successive order, which include causation as a special case, and vividness and frequency, which include resemblance as a special case. He went further than Hartley in considering the relation of sensations to the external world; external objects for him are “clusters of sensations.” Most of our beliefs about them depend on sight and sensations of color, with which we associate the other properties we attribute to them.

**CONDILLAC.** While Hartley was writing in England, Condillac was developing similar ideas in France. He was a disciple of Locke, and his first book was largely an exposition of Locke’s philosophy. In his *Traité des sensations*, he developed his own psychological theory, largely in opposition to the various current conceptions of innate ideas. He set out to show that all knowledge is “transformed sensation” and does not depend upon anything else, even, as Locke would have had it, reflection. He examined the nature and power of each of the senses by imagining a statue that has all the human faculties but has never had a sense impression. He then allowed its senses to be activated, one by one and in various combinations, and asserted that the results showed how all knowledge can gradually be constructed. He concluded that people consist of their experiences and that what they perceive is their own mental occurrences. Unlike Hartley, he did not try to give a mechanical account of these occurrences, being more concerned with psychology than physiology, and he admitted the reality of the soul. He had a considerable influence on the beginnings of British psychological thought through James Mill and J. S. Mill, Alexander Bain, and Herbert Spencer.

#### MACH AND TWENTIETH-CENTURY EMPIRICISM

Whereas Hartley, Condillac, and the Mills were interested in sensations mainly in relation to psychology, ethics and politics, Mach’s interest sprang from an attempt to provide an analysis of the methods of the physical sciences. His sensationalism was associated with a search for a solid foundation for scientific statements and with a desire to free science of all metaphysics. He held that only state-

ments which are directly verifiable in sense experience can finally be accepted as conclusions in the sciences. He concluded that all scientific statements are analyzable into statements about the relations between our sensations and that nothing can be said, scientifically, about anything beyond this. In a sentence reminiscent of James Mill he said, "The world is my sensation." It follows, also, that the various branches of science do not differ in subject matter but only in their approach to the subject matter, which is—alike for all—sensations; this was the basis of the "unity of science" movement and the logical positivism of the Vienna circle.

Mach's work was very much in harmony with the spirit of his time, especially in relation to the physical sciences, and has had an important influence on later philosophical thought. He admitted a debt to Berkeley and Hume and a number of his philosophically minded scientific contemporaries. His idea that the world is composed of "elements" which can be regarded either as sensations or the constituents of physical objects has close connections with Bertrand Russell's neutral monism and logical atomism, and his description of the aims of science is similar to that of pragmatism and operationism. In one way or another, most empiricist thought about science during the twentieth century has been influenced by his work. Recent philosophical theories of perception involving sense data or *sensa* are in the direct line of descent insofar as they stress the mind dependence of sense data, our direct awareness of or acquaintance with them, and the alleged incorrigibility of certain sorts of statements about them. Such theories can be regarded as attempts to refashion Mach's form of sensationalism in order to avoid some of the obvious objections to it.

Sensationalism and related theories all suffer from one defect, which renders the whole approach suspect; under the heading "sensations" they class together things that it is important to distinguish—for example, such sensible qualities as colors and sounds; bodily aches and pains; desires and emotions; and such feelings as dizziness, anger, and jealousy. We would not normally be prepared to class all these as experiences, but certain empiricist contentions—for example, that we know colors only through their effects on us—can make it seem superficially plausible to call them all sensations. Just because this blurs the distinctions between various things included under the heading, sensationalism as a general theory gains plausibility. Toothaches and certain feelings have an air of immediacy and unmistakability that may lead us to suppose that color sensations, since they, too, are sensations, are ultimate and incorrigible data for the

construction of a world picture. I can be certain that I have a toothache, and no one can be better justified than I in asserting or denying this. If color sensations can be assimilated to toothaches, there might seem to be some hope of arriving at incorrigible statements about the external world. Hence, the importance of the clue afforded by Anaxagoras's view that perception involves pain. A close examination of experiences of color and other sorts of experience reveals that the necessary assimilation is seriously misleading; moreover, it brings in its train enormous difficulties for an account of science. On one hand, incorrigibility can be achieved, if at all, only with the loss of the publicity of the statements concerned; on the other hand, it is difficult or impossible to show how scientific problems could ever arise if sensationalism were correct, since there is no reason that any particular combination of sensations should or should not follow any other.

In fact, the word *sensation* suffers from ambiguities similar to those involved in the word "idea" as used by Locke and Berkeley; as *sensation* must do even more work than *idea*, the ambiguities are correspondingly more serious. The view of science that springs from sensationalism, according to which science describes but does not explain, suffers further from insufficient consideration of the nature of description and its relation to explanation and from a failure to appreciate the difficulties involved in the idea of describing sensations.

*See also* Mach, Ernst; Pearson, Karl.

## Bibliography

### GENERAL

- Armstrong, D. M. *Bodily Sensations*. London: Routledge and Paul, 1962. Consideration of the logic of sensations.
- Boring, E. G. *Sensation and Perception in the History of Experimental Psychology*. New York: Appleton-Century, 1942. Useful discussions of some of the ideas of sensationalism from the point of view of psychology.
- Copleston, F. C. *History of Philosophy*. 9 vols. Westminster, MD: Newman Press, 1946–1975.
- Hamlyn, D. W. *Sensation and Perception*. New York: Humanities Press, 1961. Excellent historical survey with a useful bibliography.
- Hirst, R. J. *The Problems of Perception*. London: Allen and Unwin, 1959. Account of philosophical problems connected with sensationalism.
- Passmore, J. A. *A Hundred Years of Philosophy*. London: Duckworth, 1957.

### THE GREEKS

- Aristotle. *De Anima*. Translated by K. Foster and S. Humphries, with the commentary of St. Thomas Aquinas. London, 1951.

- Aristotle. *Works*. 12 vols, edited by J. A. Smith and W. D. Ross. Oxford, 1910–1952. See Vol. VIII, *Metaphysics*, especially  $\Gamma$ , 5 (1928), and *De Memoria et Reminiscentia*, in Vol. III (1931). These works contain most of Aristotle's important views on sensation.
- Burnet, John. *Early Greek Philosophy*. London, 1892.
- Cornford, F. M. *Plato's Theory of Knowledge*. London: K. Paul, Trench, Trubner, 1935. Contains Plato's *Theaetetus*, which considers the place of perception in knowledge, with a commentary.
- Gulley, Norman. *Plato's Theory of Knowledge*. London: Methuen, 1962. Authoritative and readable treatment.
- Kirk, G. S., and J. E. Raven. *The Presocratic Philosophers*. Cambridge, U.K.: Cambridge University Press, 1957. This and Burnet, op. cit., give accounts of pre-Socratic views on sensation.
- Plato. *The Collected Dialogues*, edited by Edith Hamilton and Huntington Cairns. New York: Pantheon, 1961. See the *Republic*, *Protagoras*, *Theaetetus*, and *Timaeus*.
- Randall, J. H. *Aristotle*. New York: Columbia University Press, 1960.

## THE SCHOLASTICS

- Copleston, F. C. *Aquinas*. Harmondsworth, U.K., Penguin, 1955. Most easily accessible account of Thomas's thought.
- Fremantle, Anne, ed. *The Age of Belief*. New York: New American Library, 1955. Selections from the writings of the Scholastics, with commentary.
- Gilson, Étienne. *Christian Philosophy in the Middle Ages*. London, 1953. Valuable general account.
- Thomas Aquinas. *Commentary on Aristotle's De Anima*. See Aristotle, *De Anima* (above).
- Thomas Aquinas. *Summa Theologiae* (selections). In *Basic Writings of St. Thomas Aquinas*, annotated by Anton C. Pegis, ed. 2 vols. New York: Random House, 1945. Vol. I, Questions 78–86.

## THE SEVENTEENTH AND EIGHTEENTH CENTURIES

- Bacon, Francis. *Novum Organum* (1620). Translated by G. W. Kitchin. Oxford, 1855.
- Bower, G. S. *Hartley and James Mill*. London, 1881.
- Condillac, É. B. de. *Essai sur l'origine des connoissances humaines*. Amsterdam, 1746.
- Condillac, É. B. de. *Traité des sensations* (1754). Translated by G. Carr as *Treatise on Sensations*. London, 1930. This and above work influenced British empiricism.
- Galileo Galilei. *Il Saggiatore* (1623). Translated in part by Stillman Drake in *Discoveries and Opinions of Galileo*. Garden City, NY: Doubleday, 1957.
- Hartley, David. *Observations on Man*. London, 1749. An early version of sensationalism that influenced both philosophers and psychologists.
- Hobbes, Thomas. *De Corpore* (1655). Translated as *Elements of Philosophy concerning Body*. London, 1839.
- Hobbes, Thomas. *Leviathan* (1651), edited by Michael Oakeshott. New York, 1947; paperback ed., 1962.
- Locke, John. *Essay concerning Human Understanding* (1690). Edited by A. C. Fraser. Oxford, 1894.
- Mill, James. *The Analysis of the Phenomena of the Human Mind*, edited by J. S. Mill. London, 1829. Development of Hartley's sensationalism.

## THE NINETEENTH AND TWENTIETH CENTURIES

- Alexander, Peter. *Sensationalism and Scientific Explanation*. London: Routledge and Paul, 1963. Criticism of the sensationalist account of science.
- Ayer, A. J. *Foundations of Empirical Knowledge*. London: Macmillan, 1940. One version of the sense-datum theory.
- Ayer, A. J., ed. *Logical Positivism*. Glencoe, IL: Free Press, 1959.
- Ayer, A. J. *Philosophical Essays*. London: Macmillan, 1954.
- Bain, Alexander. *The Senses and the Intellect*. London, 1855.
- Broad, C. D. *Scientific Thought*. London: Kegan Paul, 1923. Another modern sense-datum theory.
- Carnap, Rudolf. *The Unity of Science*. Translated by Max Black. London: Kegan Paul, 1934.
- Clifford, W. K. *Lectures and Essays*. London, 1879.
- Feigl, Herbert, and Wilfrid Sellars, eds. *Readings in Philosophical Analysis*. New York: Appleton-Century-Crofts, 1949. Useful collection of papers, many of which represent modern developments from sensationalism.
- Frank, Philipp. *Modern Science and Its Philosophy*. Cambridge, MA: Harvard University Press, 1949. Historical account of the development of logical positivism.
- George, W. H. *The Scientist in Action*. London: Williams and Norgate, 1936. Science seen as the relating of pointer readings.
- Herbart, J. F. A.B.C. *of Sense-Perception*. New York, 1896.
- Herbart, J. F. *Lehrbuch zur Psychologie*. Königsberg, Germany, 1816. Translated by M. K. Smith as *Text Book of Psychology*. London, 1891.
- James, William. *Essays in Radical Empiricism*. London: Longman, 1912.
- Kraft, Victor. *The Vienna Circle*. New York: Philosophical Library, 1953.
- Mach, Ernst. *Die Analyse der Empfindungen*. Jena, Germany, 1906. Translated by C. M. Williams as *The Analysis of Sensations*. Chicago, 1914.
- Mach, Ernst. *Populärwissenschaftliche Vorlesungen*. Leipzig, 1894. Translated by T. J. McCormack as *Popular Scientific Lectures*. Chicago: Open Court, 1897. Mach's works are classics of modern sensationalism.
- Mill, J. S. *An Examination of Sir William Hamilton's Philosophy*. London, 1865.
- Moore, G. E. *Philosophical Studies*. London: Routledge, 1922.
- Moore, G. E. *Some Main Problems of Philosophy*. London: Allen and Unwin, 1953. This and above work contain another variety of sense-datum theory.
- Neurath, Otto, Rudolf Carnap, and C. W. Morris, eds. *International Encyclopedia of Unified Science*. Chicago: University of Chicago Press, 1938–1962.
- Pearson, Karl. *The Grammar of Science*, 3rd ed. London, 1911.
- Poincaré, Henri. *La science et l'hypothèse*. Paris: Flammarion, 1902. Translated by W. J. Greenstreet as *Science and Hypothesis*. London: Walter Scott, 1905.
- Poincaré, Henri. *La valeur de science*. Paris: Flammarion, 1905. Translated by G. B. Halsted as *The Value of Science*. New York: Science Press, 1907.
- Price, H. H. *Perception*. London: Methuen, 1932. Perhaps the most detailed and elegant version of the sense-datum view.
- Russell, Bertrand. *The Analysis of Mind*. London: Macmillan, 1921.



- Russell, Bertrand. *Our Knowledge of the External World*. London: Allen and Unwin, 1914.
- Russell, Bertrand. "The Philosophy of Logical Atomism." In *Logic and Knowledge*, edited by R. C. Marsh. London: Allen and Unwin, 1956.
- Schilpp, P. A., ed. *The Philosophy of Bertrand Russell*. La Salle, IL: Open Court, 1946.
- Ward, James. *Psychological Principles*. Cambridge, U.K.: Cambridge University Press, 1918. Attack on sensationalism in psychology.

## OTHER RECOMMENDED TITLES

- Agassi, Joseph. "Sensationalism." *Mind* 75 (1966): 1–24.
- Ayers, Michael. *Locke: Epistemology and Ontology*. New York: Routledge, 1991.
- Bealer, G. "The Incoherence of Empiricism." *PAS*, Supp. 66 (1992): 99–138.
- Bennett, Jonathan. *Locke, Berkeley, Hume: Central Themes*. Oxford: Oxford University Press, 1973.
- Campbell, John. "Berkeley's Puzzle." In *Conceivability and Possibility*, edited by Tamar Szabo Gendler. Oxford: Clarendon Press, 2002.
- Chisholm, Roderick. *Perceiving: A Philosophical Study*. Ithaca, NY: Cornell University Press, 1957.
- Chisholm, Roderick. "The Problem of Empiricism." *Journal of Philosophy* 45 (1948): 512–517.
- Dicker, Georges. *Perceptual Knowledge*. Dordrecht: Reidel, 1980.
- Firth, Roderick. In *Defense of Radical Empiricism*, edited by John Troyer. Lanham, MD: Rowman & Littlefield, 1998.
- Fumerton, Richard. *Metaphysical and Epistemological Problems of Perception*. Lincoln: University of Nebraska Press, 1985.
- George, Rolf. "Kant's Sensationalism." *Synthese* 47 (1981): 229–256.
- Moser, Paul. *Knowledge and Evidence*. Cambridge, U.K.: Cambridge University Press, 1989.
- Pollock, J. *Knowledge and Justification*. Princeton, NJ: Princeton University Press, 1974.

**Peter Alexander (1967)**

*Bibliography updated by Benjamin Fiedor (2005)*

## SENSATIONS

See *Sensa*

## SENSE

"Sense" is the distinctive central notion in theories of thought and language inspired by the later work of Gottlob Frege ("sense" translates Frege's *Sinn*). For Frege what we think (not the act of thinking it) is a thought, an abstract object. Thoughts have quasi-syntactic structure. Any simple or complex constituent of a thought, even the thought itself, is a sense; thus, senses are abstract. Frege assumes that it is irrational to assent to a thought and

simultaneously dissent from it. Since someone misled about astronomy may rationally combine assent to the thought that Hesperus is Hesperus with dissent from the thought that Hesperus is Phosphorus, the thoughts are distinct. Although the names "Hesperus" and "Phosphorus" have the same reference, they express different sense, two modes of presentation of one planet. The role of a sense is to present the thinker with a reference—that is, something on which the truth-value (truth or falsity) of the thought depends; if the sense fails to present a reference, the thought lacks a truth-value. For Frege the truth-value of a thought is independent of where, when, and by whom it is thought. Thus, what reference a constituent sense presents is independent of when, where, and by whom it is thought. Sense determines reference, not vice versa.

Frege used his notion of sense to analyze the semantics of thought attributions in natural language, as in the sentence "Someone doubts that Hesperus is Phosphorus." On Frege's account expression within such "that" clauses refer to their customary senses. This explains the presumed failure of the inference from that sentence and "Hesperus is Phosphorus" to "Someone doubts that Hesperus is Hesperus": The two names have different references within "that" clauses, for their customary senses are different. If sense determines reference, then the sense of "Hesperus" in "Someone doubts that Hesperus is Phosphorus" defers from its sense in "Hesperus is Phosphorus," since the reference differs. By appeal to iterated attributions such as "He doubts that she doubts that Hesperus is Phosphorus," it can be argued that Frege is committed to an infinite hierarchy of senses. His account involves the assignment of senses to natural-language expressions. However, in order to understand many words (e.g., proper names and natural-kind terms), there is arguably no particular way in which one must think of their reference; they do not express senses common to all competent speakers. Fregeans therefore distinguish sense from linguistic meaning but in doing so sacrifice Frege's original account of thought attributions.

Sense must also be distinguished from linguistic meaning for context-dependent expressions such as "I." Two people may think "I am falling" and each refer to themselves, not the other. Since the references are distinct and sense determines reference, the senses are distinct, even though the mode of presentation is the same. Others cannot think the sense that one expresses with "I"; they can only think about it. Communication here does not amount to the sharing of thoughts, and "You think that I am falling" does not attribute to the hearer the

thought that the speaker expresses with “I am falling.” In contrast, the linguistic meaning of “I” is the same for everyone; it consists in the rule that each token of “I” refers to its producer. Unlike a sense, the rule determines reference only relative to context. Such cases reveal tensions within Frege’s conception of sense. Sense cannot be both what determines reference and how it is determined. Since senses can be qualitatively identical but numerically distinct, they are not purely abstract objects, if qualitatively identical purely abstract objects must be numerically identical.

Although Fregeans distinguish sense from linguistic meaning, they still treat a given speaker on a given occasion as expressing senses in words. Frege gave the impression that the sense expressed by a word was a bundle of descriptions that the speaker associated with it: the word refers to whatever best fits the descriptions. However, this descriptive model of reference has fared badly for proper names and natural-kind terms. Nondescriptive models may also allow different routes to the same reference, but that is a difference in sense only if it is a difference in presentation to the thinker.

In spite of these problems a role for something like sense remains. An account is needed of the deductions that thinkers are in a position to make. When, for example, is one in a position to deduce “Something is black and noisy” from “That is black” and “That is noisy”? It is necessary but not sufficient that the two tokens of “that” refer to the same thing, for, even if they do, the thinker may lack evidence to that effect: Perhaps one refers through sight, the other through hearing. What is needed is more like identity of sense than identity of reference. Thus, the theory of rational inference may still require a notion of sense. It does not follow that thinkers are always in a position to know whether given senses are identical, for it is not obvious that they are always in a position to know what deductions they are in a position to make.

**See also** Frege, Gottlob; Proper Names and Descriptions; Reference; Semantics.

### **Bibliography**

- Burge, T. “Sinning against Frege.” *Philosophical Review* 88 (1979): 398–432.
- Campbell, J. “Is Sense Transparent?” *Aristotelian Society Proceedings* 88 (1987–1988): 273–292.
- Dummett, M. *Frege: Philosophy of Language*. 2nd ed. Cambridge, MA: Harvard University Press, 1981.
- Forbes, G. “The Indispensability of *Sinn*.” *Philosophical Review* 99 (1990): 535–563.

Kripke, S. *Naming and Necessity*. Cambridge, MA: Harvard University Press, 1980.

McDowell, J. “On the Sense and Reference of a Proper Name.” *Mind* 86 (1977): 159–185.

Peacocke, C. “Sense and Justification.” *Mind* 101 (1992): 793–816.

Perry, J. *The Problem of the Essential Indexical and Other Essays*. New York: Oxford University Press, 1993.

Salmon, N. *Frege’s Puzzle*. Cambridge, MA: MIT Press, 1986.

*Timothy Williamson (1996)*

## SENSE DATA

See *Sensa*

## SENSIBILIA

See *Phenomenalism*

## SERVETUS, MICHAEL

(1511–1553)

Michael Servetus, the Spanish theologian and physician, was born in Spanish Navarre and was burned at the stake in Geneva. In the history of medicine he is remembered for having been the first to publish a description of the pulmonary circulation of the blood, and in the history of theology, he is noted for his systematic refutation of the Nicene doctrine of the Trinity. In philosophy, he developed a Christocentric pantheism that included elements from the Neoplatonic, Franciscan, and kabbalistic traditions. It should be pointed out, however, that he believed that natural philosophy should be grounded in empirical investigation.

After studying the three biblical languages as well as mathematics, philosophy, theology, and law at the universities of Zaragoza and Toulouse, Servetus, in the capacity of secretary, accompanied Juan de Quintana, the Franciscan confessor of Emperor Charles V, to the latter’s coronation in Bologna. Breaking with the imperial court, he went on his own to Basel, where he sought out John Oecolampadius, and then went on to Strasbourg, where he had some contact with Martin Bucer and, in particular, Wolfgang Capito. In nearby Hagenau he had printed his *De Trinitatis Erroribus* (1531) and, in response to Bucer’s critique, the more moderate and more Christologically oriented *De Trinitate* (1532). In Strasbourg Servetus met Kaspar Schwenkfeld, from whom he may

have taken over a heretical idea about the celestial flesh of Christ. In Strasbourg he may also have come in contact with the Anabaptists, whose views on baptism he was later to espouse. By way of Basel, where he tried to get Desiderius Erasmus's approval of his *De Trinitate*, he went to Lyons, where he worked as a proofreader and began his study of medicine under the Neoplatonizing Symphorien Champier. Next he went to Paris, where by chance he met John Calvin and got into trouble with the medical faculty over his views on astral influences. His *Apologetica Disceptatio pro Astrologia* (1538) marks an important turning point in Servetus's evaluation of the place of Greek philosophy. Whereas before he had regarded the influence of philosophy on theology as corrupting, he was now prepared to speak of "*divinus Plato*," on whose authority he defended astrology. After establishing himself in Vienne as physician to the archbishop, he engaged in correspondence with Calvin and composed the recently discovered and identified *Declaratio Jesu Christi Filii Dei* (c. 1540). Out of this grew his more massive *Restitutio Christianismi* (1553). Through the machinations of Calvin himself, Servetus was apprehended and tried for heresy, first in Catholic Lyons and then, after his escape, in Calvinist Geneva, where, after refusing to recant, he was burned at the stake.

Servetus's view of nature, history, and salvation was centered on the figure of Jesus Christ, whom he considered to be in a quite physical sense the Son of God. Servetus declined, however, to call the earthly Son eternal and declined to call either the Word or the Spirit *personae*; rather, he called them, neutrally, *res*—that is, in a modalist sense, the faces, forms, images, or manifestations of God. He mistakenly regarded the traditional *hypostasis* (*persona*) and *substantia* as equivalent, and hence, to avoid what he considered an unbiblical tritheism, he called the Father or Jehovah alone God. Before the Incarnation the Word was Elohim, or Uncreated Light. Indeed, this Light, or alternatively Christ (as distinguished from the earthly Son, Jesus), was also "the eternal sea (*pelagus*) of ideas." The Spirit has always been a Power of God, working outwardly in the world as his breath (*flatus*) and inwardly as the agitation, or motion, of the human spirit at regeneration.

The way in which the Uncreated Light became the Second Adam in Mary was for Servetus paradigmatic of the process by which creative Light was ever penetrating matter to form minerals, plants, animals, and all created things. For Servetus "even the treasures of natural science are hidden in Christ." Connected with his speculation on Light was Servetus's concept of the Shadow, according to

which he was able to regard all of the Old Testament and all religion outside the Bible as a shadowing forth of the Son that was to be born of Mary. He cherished the old Law as a pregnant woman bearing the embryonic Christ until the fullness of time.

Servetus rejected post-Constantinian (post-Nicene) Catholicism because of its alleged tritheism and its use of political force in the realm of conscience. He also opposed the Reformation churches because of their use of force, their denial of free will in accepting redemptive grace, and their neglect of sanctification, which he understood as communicated in an almost physical sense through the believers' baptism at the age of thirty (in imitation of Jesus). Nevertheless, in common with the Spiritual Libertines and some Anabaptists, Servetus held to the provisional death of every soul with the body pending the general resurrection. Under the influence of Joachimite speculation, he believed that the true church would be restored in the year 1560.

### Bibliography

#### WORKS BY SERVETUS

The two works on the Trinity have been translated by E. M. Wilbur as *Two Treatises of Servetus on the Trinity*, Harvard Theological Studies (Cambridge, MA: Harvard University Press, 1932), Vol. 16. See also Charles D. O'Malley, *Michael Servetus: A Translation of His Geographical, Medical and Astrological Writings* (Philadelphia: American Philosophical Society, 1953).

#### WORKS ON SERVETUS

For literature on Servetus, see Roland Bainton, *Hunted Heretic: The Life and Death of Michael Servetus, 1511–1553* (Boston: Beacon Press, 1953), which includes a bibliography; Wolrad Emde, "Michael Servetus als Renaissancephilosoph und Restitutionstheologe," in *Zeitschrift für Kirchengeschichte* 50 (1941): 96–131, especially good on *Restitutio Christianismi*; and Marcial Solana, *Historia de la filosofía española* (Madrid: Real Academia de Ciencias Exactas, Físicas y Naturales, 1941), Vol. I, pp. 629–681, especially good on *De Trinitate*.

George Hunston Williams (1967)

## SETH, ANDREW

See Pringle-Pattison, Andrew Seth

## SET THEORY

Set theory is a mathematical theory of collections, "sets," and collecting, as governed by axioms. Part of its larger significance is that mathematics can be reduced to set

theory, with sets doing the work of mathematical objects and their collections and set-theoretic axioms providing the basis for mathematical proofs. With this reduction in play, modern set theory has become an autonomous and sophisticated research field of mathematics, enormously successful at the continuing development of its historical heritage as well as at analyzing strong propositions and gauging their consistency strength.

Set theory arose in mathematics in the late nineteenth century as a theory of infinite collections and soon became intertwined with the development of analytic philosophy and mathematical logic. The subject was then developed as the logical distinction was being clarified between “falling under a concept,” to be transmuted in set theory to “ $x \in y$ ,”  $x$  is a member of  $y$ , and subordination or inclusion, to be transmuted in set theory to “ $x \subseteq y$ ,”  $x$  is a subset of  $y$ . That set theory is both a field of mathematics and serves as a foundation for mathematics emerged early in this development.

In what follows, set theory is presented as both a historical as well as an epistemological phenomenon, driven forward by mathematical problems, arguments, and procedures. The first part describes the groundbreaking work of Georg Cantor on infinite sets analyzed in terms of power, transfinite numbers, and well-orderings. The next two parts describe the subsequent transmutation of the notion of set through axiomatization, a process to be associated largely with Ernst Zermelo. Next will come a description of the work of Kurt Gödel on the constructible sets, work that made first-order logic central to set theory, followed by a description of the work of Paul Cohen on forcing, a method that transformed set theory into a modern, sophisticated field of mathematics. The last section describes the modern investigation of relative consistency in terms of forcing, large cardinals, and inner models.

## POWER, NUMBER, AND WELL-ORDERING

Set theory was born on that day in December 1873 when Cantor established that *the continuum is not countable*. The concepts here are fundamental: Taking infinite collections as unitary totalities, a set is *countable* if it is in one-to-one correspondence with the set of natural numbers  $\{0, 1, 2, \dots\}$ , and the *continuum* is the linear continuum regarded extensionally as a collection of points corresponding to the real numbers. In a 1878 publication Cantor investigated ways of defining one-to-one correspondences between sets. For sets of real numbers and the like, he stipulated that two sets have the same *power* if

there is a one-to-one correspondence between them and that a set  $x$  has a *higher* power than a set  $y$  if  $y$  has the same power as a subset of  $x$  yet  $x$  and  $y$  do not have the same power. He managed to show that the continuum, the plane, and generally  $n$ -dimensional Euclidean space all have the same power, but at this point in mathematics there were still only the two infinite powers as set out by his 1873 proof. Cantor at the end of his 1878 publication conjectured:

Every infinite set of real numbers either is countable or has the power of the continuum.

This was the *Continuum Hypothesis* (CH) in its nascent context. The *Continuum Problem* would be to resolve this hypothesis, and viewed as a primordial problem it would stimulate Cantor both to approach the real numbers in an increasingly arithmetical fashion and to grapple with fundamental questions of set existence.

In his magisterial *Grundlagen* of 1883, Cantor developed the *transfinite numbers* and the key concept of *well-ordering*. Investing the “symbols of infinity” of his early trigonometric series investigations with a new autonomy, Cantor conceived of the transfinite numbers as being generated by the operations of taking successors and of taking limits of increasing sequences. Extending beyond the finite  $0, 1, 2, \dots$ , the progression of transfinite numbers could be depicted, in his later notation, in terms of natural extensions of arithmetical operations:

$$0, 1, 2, \dots, \omega, \omega + 1, \omega + 2, \dots, \omega + \omega (= \omega \cdot 2), \dots, \omega \cdot 3, \dots, \omega \cdot \omega (= \omega^2), \dots, \omega^3, \dots, \omega^\omega, \dots$$

*Definition.* A binary relation  $<$  is a *linear ordering* of a set  $a$  if it is transitive, that is,  $x < y$  and  $y < z$  implies  $x < z$ , and trichotomous, that is, for  $x, y \in a$ , exactly one of  $x < y$ ,  $x = y$ , or  $y < x$  holds.

A relation  $<$  is a *well-ordering* of a set  $a$  if it is a linear ordering of the set such that every non-empty subset has a  $<$ -least element.

Well-orderings convey the sense of sequential counting, and the transfinite numbers serve as standards for gauging well-orderings. As Cantor pointed out, every linear ordering of a finite set is already a well-ordering and all such orderings are isomorphic, so that the general sense is only brought out by infinite sets. For these there could be non-isomorphic well-orderings. For example the set of natural numbers  $\{0, 1, 2, \dots\}$ , that is, the predecessors of  $\omega$ , can be put into one-to-one correspondence with the predecessors of  $\omega + \omega$  by sequentially counting the evens before the odds. In fact all the infinite transfinite numbers in the above display are countable. Cantor

called the set of natural numbers the first number class (I) and the set of numbers whose predecessors are in one-to-one correspondence with (I) the second number class (II). Cantor conceived of (II) as bounded above according to a limitation principle and showed that (II) itself is not countable. Proceeding upward, Cantor called the set of numbers whose predecessors are in one-to-one correspondence with (II) the third number class (III), and so forth. In this way Cantor conceived of ever higher powers as represented by number classes and moreover took every power to be so represented. With this “free creation” of numbers, Cantor then propounded in section 3 of the *Grundlagen* a basic principle that was to drive the analysis of sets:

It is always possible to bring any *well-defined set* into the form of a *well-ordered set*.

He regarded this as a “an especially remarkable law of thought which through its general validity is fundamental and rich in consequences.” Sets are to be well-ordered and thus to be gauged via the transfinite numbers of his structured conception of the infinite.

The transfinite numbers provided the framework for Cantor’s two approaches to the Continuum Problem, one through power and the other through definable sets of real numbers, these each to initiate two vast research programs. As for the first, Cantor in the *Grundlagen* established results that reduced the Continuum Problem to showing that the continuum and the second number class have the same power. However, despite several announcements Cantor could never develop a workable correlation, an emerging problem being that he could not define a well-ordering of the real numbers. As for the approach through definable sets of real numbers, Cantor showed that “CH holds for closed sets.” Closed sets are a very simple kind of definable set of real numbers, and Cantor showed that a closed set either is countable or has the power of the continuum. He thus reduced the Continuum Problem to determining whether there is a closed set of real numbers of the power of the second number class. He could not do this, but he had established the first result of *descriptive set theory*, the definability theory for the continuum.

Almost two decades after his initial 1873 proof, Cantor in a short 1891 note gave his now celebrated *diagonal* argument. He proceeded in terms of functions, ushering in collections of arbitrary functions into mathematics, but we state and prove his result as is done nowadays in terms of the power set  $\mathcal{P}(x) = \{y \mid y \subseteq x\}$  of a set  $x$ , the collection of all its subsets: *For any set  $x$ ,  $\mathcal{P}(x)$  has a higher power than  $x$ .*

First, the function associating each  $a \in x$  with  $\{a\}$ , that subset of  $x$  with sole member  $a$ , is a one-to-one correspondence between  $x$  and a subset of  $\mathcal{P}(x)$ . Assume now to the contrary that there is a one-to-one correspondence  $F$  established between the members of  $x$  and all the members of  $\mathcal{P}(x)$ . Consider the “diagonal” set  $d = \{a \mid a \in x \text{ and } a \notin F(a)\}$  consisting of those members  $a$  of  $x$  that do not belong to their corresponding subset  $F(a)$ . If  $d$  itself were a value of  $F$ , say  $d = F(b)$  for some  $b \in x$ , then we would have the paradigmatic contradiction:  $b \in d$  exactly when  $b \notin d$ . Hence,  $F$  was not a one-to-one correspondence after all!

Cantor had been shifting his notion of set to a level of abstraction beyond sets of real numbers and the like, and the casualness of his 1891 note may reflect an underlying cohesion with his earlier 1873 argument. Indeed the diagonal argument can be drawn out of the earlier argument, and the new result generalized the old since, with  $N$  the set of natural numbers,  $\mathcal{P}(N)$  is in one-to-one correspondence with the continuum. With his new result Cantor affirmed that the powers of well-defined sets have no maximum, and he had proved for the first time that there is a power greater than that of the continuum. However, with his view that every well-defined set is well-ordered Cantor would now have had to confront, in his arbitrary function context, a general difficulty starkly abstracted from the Continuum Problem: *From a well-ordering of a set a well-ordering of its power set is not necessarily definable.* The diagonal proof called into question Cantor’s very notion of set.

Cantor’s *Beiträge*, published in two parts in 1895 and 1897, presented his mature theory of the transfinite. In the first part Cantor reconstrued power as *cardinal number*, an autonomous concept beyond being *une façon de parler* about one-to-one correspondence. He defined the addition, multiplication, and exponentiation of cardinal numbers primordially in terms of set-theoretic operations and functions. As befits the introduction of new numbers Cantor then introduced a new notation, one using the Hebrew letter aleph,  $\aleph$ . With  $\aleph_0$  the cardinal number of the set of natural numbers Cantor showed that

$$\aleph_0 \cdot \aleph_0 = \aleph_0 \text{ and } 2^{\aleph_0} \text{ is the cardinal number of the continuum}$$

(and hence of  $\mathcal{P}(N)$ ). With this he observed that the 1878 labor of associating the continuum with the plane and so forth could be reduced to a “few strokes of the pen” in his new arithmetic. Cantor only mentioned

$$\aleph_0, \aleph_1, \aleph_2, \dots, \aleph_\omega, \dots,$$

these to be the cardinal numbers of the successive number classes from the *Grundlagen* and thus to exhaust all the infinite cardinal numbers.

Cantor then developed his theory of *order types*, “types” or abstractions of linear orderings. He defined the addition and multiplication of order types and characterized the order types of the rational numbers and of the real numbers. In the second *Beiträge* Cantor turned to the special case of well-orderings and reconstrued the transfinite numbers as their order types, newly calling the numbers the *ordinal numbers*. He then established their basic comparability properties by showing that given two well-orderings one is isomorphic to an initial segment of the other or vice versa. In this new setting he concentrated on the countable ordinal numbers, the new construal of the second number class, and provided an incisive structural analysis in terms of a new operation of ordinal exponentiation.

The two parts of the *Beiträge* were not only distinct by subject matter, cardinal number and the continuum vs. ordinal number and well-ordering, but also between them there developed a wide, insurmountable breach. In the first part nowhere is the 1891 result stated even in a special case, though it was now possible to express it as  $\mathfrak{m} < 2^{\mathfrak{m}}$  for any cardinal number  $\mathfrak{m}$ , since in his arithmetic

$2^{\mathfrak{m}}$  is the cardinal number of the power set of a set with cardinal number  $\mathfrak{m}$ .

Also, the second *Beiträge* does not mention any aleph beyond  $\aleph_1$ , nor does it mention the Continuum Hypothesis, which could have been stated as  $2^{\aleph_0} = \aleph_1$ . Every well-ordered set, through a corresponding ordinal number, has an aleph as its cardinal number, but how does  $2^{\aleph_0}$  fit into the aleph sequence?

Thus the Continuum Problem was embedded in the very interstices of the early development of set theory, and in fact the structures that Cantor built, while now of great intrinsic interest, emerged out of efforts to articulate and solve the Continuum Problem. The tension uncovered by Cantor’s diagonal argument between well-ordering and power set (or arbitrary functions) would soon be revisited by Zermelo. David Hilbert, when he presented his famous list of twenty-three problems at the 1900 International Congress of Mathematicians at Paris, made the Continuum Problem the very first problem and intimated Cantor’s difficulty by suggesting the desirability of “actually giving” a well-ordering of the real numbers.

At the turn into the twentieth century the “logical” limits of set formation and existence were broached for sets being counterparts to “concepts” or properties. In correspondence with Hilbert and Richard Dedekind in the late 1890s Cantor became newly engaged with questions of set existence. He had earlier considered collections like all ordinal numbers or all alephs as leading out of his conceptual framework. These “absolutely infinite or inconsistent multiplicities,” if admitted as sets, would lead to contradictions, and Cantor argued anew that every set can be well-ordered else it would in one-to-one correspondence with all the ordinal numbers and hence an inconsistent multiplicity. In this he anticipated later developments in set theory.

Bertrand Russell, a main architect of the analytic tradition in philosophy, focused in 1900 on Cantor’s work. Russell was pivoting from idealism toward logicism, the thesis that mathematics can be founded in logic. Taking a universalist approach to logic with all-encompassing categories, Russell took the class of all classes to have the largest cardinal number but saw that Cantor’s 1891 result leading to higher cardinal numbers presented a problem. Analyzing that argument, by the spring of 1901 he arrived at the famous *Russell’s Paradox*. This paradox showed with remarkable simplicity that there are properties  $P(x)$  such that the collection of objects having that property, the *class*

$$\{x \mid P(x)\},$$

cannot itself be an object: Consider  $\{x \mid x \notin x\}$ . If this were an object  $r$  in the range of possibilities, then we would have the contradiction  $r \in r$  exactly when  $r \notin r$ . This paradox may have been critical for Russell’s universalist approach to logic and for logicism, but it was less so for the development of set theory, which was emerging in mathematics. In any case the paradox did serve as a motivation for fashioning a consistent notion of set through axiomatization.

The first decade of the new century saw Zermelo make his major advances in the development of set theory. Already estimable as an applied mathematician, Zermelo turned to set theory and its foundations under the influence of Hilbert. Zermelo’s first substantial result was his independent discovery of the argument for Russell’s Paradox. He then established in 1904 the Well-Ordering Theorem, that every set can be well-ordered, assuming what he soon called the Axiom of Choice (AC). Zermelo thereby shifted the notion of set away from the implicit assumption of Cantor’s principle that every well-defined

set is well-ordered and replaced that principle by an explicit axiom about a wider notion of set.

In retrospect Zermelo's *argument* for his Well-Ordering Theorem can be viewed as pivotal for the development of set theory. To summarize the argument, suppose that  $x$  is a set to be well-ordered, and through Zermelo's Axiom-of-Choice hypothesis assume that the power set  $\mathcal{P}(x) = \{y \mid y \subseteq x\}$  has a choice function, that is, a function  $\gamma$  such that for every non-empty member  $y$  of  $\mathcal{P}(x)$ ,  $\gamma(y) \in y$ . Call a subset  $y$  of  $x$  a  $\gamma$ -set if there is a well-ordering  $R$  of  $y$  such that for each  $a \in y$ ,

$$\gamma(\{z \mid z \notin y \text{ or } z R a \text{ fails}\}) = a.$$

That is, each member of  $y$  is what  $\gamma$  "chooses" from what does not already precede that member according to  $R$ . The main observation is that  $\gamma$ -sets cohere in the following sense: If  $y$  is a  $\gamma$ -set with well-ordering  $R$  and  $z$  is a  $\gamma$ -set with well-ordering  $S$ , then  $y \subseteq z$  and  $S$  is a prolongation of  $R$ , or vice versa. With this, let  $w$  be the union of all the  $\gamma$ -sets, that is, all the  $\gamma$ -sets put together. Then  $w$  too is a  $\gamma$ -set, and by its maximality it must be all of  $x$  and hence  $x$  is well-ordered.

Note that the converse to this result is immediate in that if  $x$  is well-ordered, say with a well-ordering  $<$ , then the power set  $\mathcal{P}(x)$  has a choice function  $\delta$ , namely for each non-empty member  $y$  of  $\mathcal{P}(x)$ , let  $\delta(y)$  be the the  $<$ -least member of  $y$ . Not only did Zermelo's argument analyze the connection between well-ordering and choice functions, but it anticipated in its defining of approximations and taking of a union the proof procedure for von Neumann's Transfinite Recursion Theorem.

Zermelo maintained that the Axiom of Choice, to the effect that *every* set has a choice function, is a "logical principle" which "is applied without hesitation everywhere in mathematical deduction," and this is reflected in the Well-Ordering Theorem being regarded as a theorem. Cantor's work had served to exacerbate a growing discord among mathematicians with respect to two related issues: whether infinite collections can be mathematically investigated at all, and how far the function concept is to be extended. The positive use of an arbitrary function operating on arbitrary subsets of a set having been made explicit, there was open controversy after the appearance of Zermelo's proof. This can be viewed as a turning point for mathematics, with the subsequent tilting toward the acceptance of the Axiom of Choice symptomatic of a conceptual shift in mathematics.

## AXIOMATIZATION

In response to his critics Zermelo published a second proof of the Well-Ordering Theorem in 1908, and with axiomatization assuming a general methodological role in mathematics he also published in 1908 the first full-fledged axiomatization of set theory. But as with Cantor's work, this was no idle structure building but a response to pressure for a new mathematical context. In this case it was not for the formulation and solution of a *problem* like the Continuum Problem, but rather to clarify a *proof*. Zermelo's motive in large part for axiomatizing set theory was to buttress his Well-Ordering Theorem by making explicit its underlying set existence assumptions. Effecting the first transmutation of the notion of set after Cantor, Zermelo ushered in a new abstract, prescriptive view of sets as solely structured by membership and governed by axioms, a view that would soon come to dominate.

The following are Zermelo's axioms, much as they would be presented today. They are to govern the connections between  $\in$  and  $\subseteq$  and to prescribe the generation of new sets out of old. The standard axiomatization would be the result of adding two further axioms and formalizing in first-order logic.

**AXIOM OF EXTENSIONALITY.** *Two sets are equal exactly when they have the same members.* Thus sets epitomize the *extensional* view of mathematics, it being stipulated that however sets are arrived at, there is a definite criterion for equality provided solely by membership.

**AXIOM OF EMPTY SET.** *There is a set having no members.* This axiom serves to emphasize the beginning with an initial set, the *empty* set, denoted  $\emptyset$ .

**AXIOM OF PAIRS.** *For any sets  $x$  and  $y$ , there is a set consisting of exactly  $x$  and  $y$  as members.* The posited set is denoted  $\{x,y\}$  and is called the (unordered) pair of  $x$  and  $y$ .  $\{x,x\}$  is denoted  $\{x\}$ , as we have already seen, and is called the *singleton* of  $x$ .

**AXIOM OF UNION.** *For any set  $x$ , there is a set consisting exactly of those sets that are members of some member of  $x$ .* The posited set is denoted  $\cup x$  and is called the *union* of  $x$ . This "generalized" union subsumes the better known binary union, in that for any sets  $a$  and  $b$ ,

$$a \cup b = \cup \{a,b\} = \{x \mid x \in a \text{ or } x \in b\}.$$

If a set  $x$  is structured as an indexed set  $\{x_i \mid i \in I\}$ , then  $\cup x$  is often written as  $\cup_{i \in I} x_i$  or just  $\cup x_i$ .

**AXIOM OF POWER SET.** For any set  $x$ , there is a set consisting exactly of the subsets of  $x$ . The posited set is denoted  $\mathcal{P}(x)$  and is called the *power set* of  $x$ , as we have already seen.

**AXIOM OF CHOICE.** For any set  $x$  consisting of non-empty, pairwise disjoint sets, there is a set  $c$  such that every member of  $x$  has exactly one element in  $c$ . Thus,  $c$  acts like a choice function for  $x$  construed as a family of sets. This is a reductive way of positing choice functions.

**AXIOM OF INFINITY.** There is a set having  $\emptyset$  as a member and such that whenever  $y$  is a member, so also is  $y \cup \{y\}$ . This has become the usual way of positing the existence of an infinite set, in light of the definition of ordinals. Zermelo actually stated his axiom with “ $y \cup \{y\}$ ” replaced by “ $\{y\}$ ,” getting at a set describable informally as  $\{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \dots\}$ .

**AXIOM OF SEPARATION.** For any set  $x$  and definite property  $P$ , there is a set consisting exactly of those members of  $x$  having the property  $P$ . Once a collection has been comprehended as a set, we are able to form a subset by “separating” out according to a property. Or, a subclass of a set is a set. Taking the property of being a member of a given set  $a$ , we have as a set the binary intersection

$$x \cap a = \{y \mid y \in x \text{ and } y \in a\}.$$

Taking the property of not being a member of  $a$ , we have as a set the set-theoretic difference

$$x - a = \{y \mid y \in x \text{ and } y \notin a\}.$$

As a further use of the axiom, consider for a set  $x$  the *intersection* of  $x$ :

$$\cap x = \{a \mid a \in y \text{ for every } y \in x\}.$$

This is a (property-specifiable) subclass of any member of  $x$ , and so we have as a theorem: *If  $x \neq \emptyset$  then  $\cap x$  is a set.* This is a “generalized” intersection, with the better known binary intersection being  $\cap \{x, a\} = x \cap a$ .

According to Zermelo a property is “definite if the fundamental relations of the domain, by means of the axioms and the universally valid laws of logic, determine without arbitrariness whether it holds or not.” But with no underlying logic formalized, the ambiguity of definite property would become a major issue, one that would eventually be resolved only decades later through first-order formalization. In any case Zermelo saw that the Separation idea suffices for a development of set theory that still allows for the “logical” formation of sets accord-

ing to property. Russell’s Paradox is forestalled since only “logical” subsets are to be allowed; indeed, Zermelo’s first theorem was that there is no universal set, a set that contains every set as a member, the *reductio* argument being the paradox argument.

Stepping back, Extensionality, Empty Set, and Pairs served to lay the basis for sets. Infinity and Power Set ensured sufficiently rich settings for set-theoretic constructions. Tempering the logicians’ extravagant and problematic “all,” Power Set provided the provenance for “all” for subsets of a given set, just as Separation served to capture “all” for elements of a given set satisfying a property. Finally, Union and Choice completed the encasing of Zermelo’s proof(s) of his Well-Ordering Theorem in the necessary set existence principles.

Although Hilbert’s axiomatization of geometry in his 1899 *Grundlagen der Geometrie* may have served as a model for Zermelo’s axiomatization of set theory and Dedekind’s 1888 essay *Was sind und was sollen die Zahlen?* on the foundations of arithmetic a precursor, there are crucial differences having to do with subject matter and proof. Both in intent and outcome Dedekind and Hilbert had been engaged in the analysis of fixed subject matter. Dedekind in particular had done a great deal to enshrine proof as the vehicle for algebraic abstraction and generalization. Like algebraic constructs, sets were new to mathematics and would be incorporated by setting down rules for their proofs. Just as Euclid’s axioms for geometry had set out the permissible geometric constructions, the axioms of set theory would set out rules for set generation and manipulation. But unlike the emergence of mathematics from marketplace arithmetic and Greek geometry, sets and transfinite numbers were neither laden with nor bolstered by substantial antecedents. There was no fixed, intended subject matter. Like strangers in a strange land stalwarts developed a familiarity with sets guided step by step by the axiomatic framework. For Dedekind it had sufficed to work with sets by merely giving a few definitions and properties, those foreshadowing Extensionality, Union, and Infinity. Zermelo provided more rules: Separation, Power Set, and Choice.

Zermelo’s 1908 axiomatization paper, especially with its rendition at the end of the Cantorian theory of cardinality in terms of functions cast as set constructs, brought out Zermelo’s *set-theoretic reductionism*. Zermelo pioneered the reduction of mathematical concepts and arguments to set-theoretic concepts and arguments from axioms, based on sets doing the work of mathematical objects. Set theory would provide the underpinnings of mathematics, and Zermelo’s axioms would resonate with



emerging mathematical practice. Zermelo's analysis moreover served to draw what would come to be generally regarded as set-theoretic out of the realm of the presumptively logical. This would be particularly salient for Infinity and Power Set and was strategically advanced by the segregation of property considerations to Separation. Based on generative and prescriptive axioms, set theory would become more combinatorial, less logical. With these features Zermelo's axioms indeed proved more than adequate to serve as a reductive basis for mathematics, at least for providing surrogates for mathematical objects; looking ahead it was for subsequent developments to bring out that set theory could also serve as a court of adjudication in terms of relative consistency.

Felix Hausdorff was the first developer of the transfinite after Cantor, the one whose work first suggested the rich possibilities for a mathematical investigation of the higher transfinite. A mathematician *par excellence* Hausdorff took the sort of mathematical approach to set theory and set-theoretic approach to mathematics which would come to dominate in the years to come. In a 1908 publication Hausdorff brought together his extensive work on *uncountable* order types, and in particular formulated the *Generalized Continuum Hypothesis* (GCH): For any infinite set  $x$ , there is no set of cardinal number strictly intervening between that of  $x$  and of its power set  $\mathcal{P}(x)$ ; or in Cantor's later terms, for every ordinal number  $\alpha$ ,  $2^{\aleph_\alpha} = \aleph_{\alpha+1}$ . Hausdorff also entertained for the first time a "large cardinal" concept, of which more below. Hausdorff's classic 1914 text, *Grundzüge der Mengenlehre*, broke the ground for a generation of mathematicians in both set theory and topology. He presented Cantor's and Zermelo's work systematically, and of particular interest, he applied the Axiom of Choice to provide what is now known as Hausdorff's Paradox. The source of the later and better known Banach-Tarski Paradox, Hausdorff's Paradox provided an implausible decomposition of the sphere and was the first, and a dramatic, synthesis of classical mathematics and the new Zermelian abstract view.

In the *Grundzüge* Hausdorff defined an ordered pair of sets in terms of (unordered) pairs, formulated functions in terms of ordered pairs, and ordering relations as collections of ordered pairs. Hausdorff thus capped efforts of logicians by making their moves in mathematics, completing the set-theoretic reduction of relations and functions. In the modern setting, the definition of the ordered pair that has been adopted is not Hausdorff's, but one provided by Kazimierz Kuratowski in 1921:

$$\langle x, y \rangle = \{\{x\}, \{x, y\}\}.$$

This satisfies all that is operationally required of an ordered pair:

$$\langle x, y \rangle = \langle a, b \rangle \text{ exactly when } x = a \text{ and } y = b.$$

With this definition, a set  $r$  is a *relation* if it consists of ordered pairs. This objectification is often eased by reverting to the older conceptual notation  $a r b$  for  $\langle a, b \rangle \in r$ . A set  $f$  is a *function* if it is a relation satisfying: If  $\langle x, y \rangle \in f$  and  $\langle x, z \rangle \in f$ , then  $y = z$ . This objectification is eased by reverting to the older operational notation  $f(x) = y$  for  $\langle x, y \rangle \in f$ , though the emphasis is on the generality and arbitrariness of  $f$  as just a relation with a univalency property. Finally the dynamic notation  $f: a \rightarrow b$  specifies that  $f$  is a function such that every member of  $a$  is a first coordinate of an ordered pair in  $f$ , and that every second coordinate is a member of  $b$ .

### AXIOMATIZATION COMPLETED

In the 1920s fresh initiatives structured the loose Zermelian framework with new features and corresponding developments in axiomatics, the most consequential moves made by John von Neumann with anticipations by Dimitry Mirimanoff in a pre-axiomatic setting. Von Neumann effected a Counter-Reformation of sorts that led to the incorporation of a new axiom, the Axiom of Replacement: The transfinite numbers had been central for Cantor but peripheral to Zermelo; von Neumann reconstrued them as *bona fide* sets, the ordinals, and established their efficacy by formalizing transfinite recursion, the method of sequential definition of sets based on previously defined sets applied with transfinite indexing.

Ordinals manifest the basic idea of taking precedence in a well-ordering simply to be membership:

*Definitions.* A set  $x$  is *transitive* if  $\cup x \subseteq x$ , that is, whenever  $a \in b$  and  $b \in x$ , then  $a \in x$ .

A set  $x$  is a (von Neumann) *ordinal* if  $x$  is transitive and the membership relation restricted to  $x = \{y \mid y \in x\}$  is a well-ordering of  $x$ .

For example,  $\emptyset$  is transitive, but  $\{\{\emptyset\}\}$  is not. Loosely speaking, transitive sets retain all their hereditary members. The first several ordinals are

$$\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset, \{\emptyset, \{\emptyset\}\}\}, \dots$$

and are newly taken to be the numbers 0, 1, 2, 3, ... . If  $x$  is an ordinal, then so also is  $x \cup \{x\}$ , the *successor* of  $x$ , and this accounts for how the Axiom of Infinity was formulated in the previous section. It has become customary to use the Greek letters  $\alpha, \beta, \gamma, \dots$  to denote ordinals. Von

Neumann, as had Mirimanoff before him, established the key instrumental property of Cantor's ordinal numbers for ordinals: *Every well-ordered set is order-isomorphic to exactly one ordinal with membership*. The proof made a paradigmatic use of Replacement, and so was the first proof to draw that axiom into set theory.

For a set  $x$  and property  $P(v, w)$ , the property is said to be *functional on  $x$*  if for any  $a \in x$ , there is exactly one  $b$  such that  $P(a, b)$ .

**AXIOM OF REPLACEMENT.** For any set  $x$  and property  $P(v, w)$  functional on  $x$ ,  $\{b \mid P(a, b) \text{ for some } a \in x\}$  is a set.

This axiom allows for new sets that result when members of a set are “replaced” according to a property. If the functional property is given by a set, that is there is a function  $f$ , a set of ordered pairs, such that  $P(v, w)$  exactly when  $f(v) = w$ , then Replacement is not needed. However, as in the case of the above-stated result correlating arbitrary well-orderings with ordinals, there are functional properties that are more general, typically formulated by recursion.

Replacement subsumes Separation. Suppose that  $x$  is a set and  $P$  is a (definite) property. If there are no members of  $x$  satisfying  $P$ , then we are done. Otherwise, fix such a member  $y_0$ . For any  $a \in x$ , let  $P(a, a)$  hold if  $a$  satisfies  $P$  and  $P(a, y_0)$  hold otherwise. Then the “replaced” set  $\{b \mid P(a, b) \text{ for some } a \in x\}$  is the set of members of  $x$  satisfying  $P$ .

Von Neumann took the crucial step of ascribing to the ordinals the role of Cantor's ordinal numbers with their several principles of generation. Now, with ordinal numbers regarded as gauging well-orderings, that one is isomorphic to a proper initial part of another corresponds for ordinals to actual membership and can be rendered

$$\alpha < \beta \text{ exactly when } \alpha \in \beta.$$

For this reconstrual of ordinal numbers and already to define the arithmetic of ordinals von Neumann saw the need to establish the Transfinite Recursion Theorem, the theorem that validates definitions by recursion along well-orderings. The proof was anticipated by the Zermelo 1904 proof, but Replacement was necessary even for the very formulation, let alone the proof, of the theorem. With the ordinals in place von Neumann completed the restoration of the Cantorian transfinite by defining the *cardinals* as the *initial ordinals*, those ordinals not in one-to-one correspondence with any of its predecessors. The infinite initial ordinals are denoted

$$\omega = \omega_0, \omega_1, \omega_2, \dots, \omega_\alpha, \dots,$$

so that  $\omega$  is to be the set of natural numbers in the ordinal construal, and the identification of different intentions is signaled by

$$\omega_\alpha = \aleph_\alpha$$

with the left being a von Neumann ordinal and the right being the Cantorian cardinal number. Every set  $x$ , with AC, is well-orderable and hence in one-to-one correspondence with an initial ordinal  $\omega_\alpha$ , and the *cardinality* of  $x$  is  $|x| = \aleph_\alpha$ . It has become customary to use the middle Greek letters  $\kappa, \lambda, \mu, \dots$  to denote initial ordinals in their role as the cardinals. A *successor* cardinal is one of form  $\aleph_{\alpha+1}$  and is denoted  $\kappa^+$  for  $\kappa = \aleph_\alpha$ . A cardinal which is not a successor cardinal is a *limit* cardinal.

Replacement has been latterly regarded as somehow less necessary or crucial than the other axioms, the purported effect of the axiom being only on large-cardinality sets. Initially Abraham Fraenkel and Thoralf Skolem had independently in 1922 proposed the addition of Replacement to Zermelo's axioms, both pointing out the inadequacy of Zermelo's axioms for establishing that  $E = \{Z_0, \mathcal{P}(Z_0), \mathcal{P}(\mathcal{P}(Z_0)), \dots\}$  is a set, where  $Z_0 = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \dots\}$  is Zermelo's infinite set from his Axiom of Infinity. However even  $F = \{\emptyset, \mathcal{P}(\emptyset), \mathcal{P}(\mathcal{P}(\emptyset)), \dots\}$  cannot be proved to be a set from Zermelo's axioms: The union of  $E$  above, with membership restricted to it, models Zermelo's axioms yet does not have  $F$  as a member. Hence Zermelo's axioms cannot establish the existence of some simple countable sets consisting of finite sets and could be viewed as remarkably lacking in closure under finite recursive processes. If the Axiom of Infinity were itself modified to entail that  $F$  is a set, then there would still be many other finite sets  $a$  so that  $\{a, \mathcal{P}(a), \mathcal{P}(\mathcal{P}(a)), \dots\}$  cannot be proved to be a set. Replacement serves to rectify the situation by allowing new infinite sets defined by “replacing” members of the one infinite set given by the Axiom of Infinity. In any case the full exercise of Replacement is part and parcel of transfinite recursion, and it was von Neumann's formal incorporation of this method into set theory, as necessitated by his proofs, that brought in Replacement.

Von Neumann (and before him Mirimanoff, Fraenkel, and Skolem) also considered the salutary effects of restricting the universe of sets to the well-founded sets. The *well-founded* sets are the sets that belong to some “rank”  $V_\alpha$ , these definable through transfinite recursion:

$$V_0 = \emptyset; V_{\alpha+1} = \mathcal{P}(V_\alpha); \text{ and } V_\delta = \bigcup \{V_\alpha \mid \alpha < \delta\} \text{ for limit ordinals } \delta.$$

$V_\omega$  consists of the “hereditarily finite” sets,  $\omega \in V_{\omega+1}$  and  $\mathcal{P}(\omega) \in V_{\omega+2}$ , and so already in these beginning levels there are set counterparts for many objects in mathematics. That the universe  $V$  of all sets is the *cumulative hierarchy*

$$V = \cup \{V_\alpha \mid \alpha \text{ is an ordinal}\}.$$

is thus the assertion that every set is well-founded. Von Neumann essentially showed that this assertion is equivalent to a simple assertion about sets:

**AXIOM OF FOUNDATION.**  $\forall x(x \neq \emptyset \rightarrow \exists y \in x (x \cap y = \emptyset)).$

Thus non-empty well-founded sets have  $\in$ -minimal members. If a set  $x$  satisfies  $x \in x$  then  $\{x\}$  is not well-founded; similarly if there are  $x_1 \in x_2 \in x_1$ , then  $\{x_1, x_2\}$  is not well-founded. Ordinals and sets consisting of ordinals are well-founded, and well-foundedness can be viewed as a generalization of being an ordinal that loosens the connection with transitivity. The Axiom of Foundation eliminates pathologies like  $x \in x$  and through the cumulative hierarchy rendition provides metaphors about building up the universe of sets and the possibility of inductive arguments to establish results about all sets.

In a remarkable 1930 publication Zermelo offered his final axiomatization of set theory as well as a striking, synthetic view of a procession of models that would have a modern resonance. Proceeding in what we would now call a second-order context, Zermelo extended his 1908 axiomatization by adjoining both Replacement and Foundation. The standard axiomatization of set theory

ZFC, *Zermelo-Fraenkel with Choice*,

is recognizable, the main difference being that ZFC is a first-order theory (see the next section); “Fraenkel” acknowledges Fraenkel’s suggestion of adjoining Replacement; and the Axiom of Choice is explicitly mentioned.

ZF, *Zermelo-Fraenkel*,

is ZFC without AC and is a base theory for the investigation of weak Choice-type propositions as well as propositions that contradict AC.

Zermelo herewith completed his transmutation of the notion of set, his abstract, prescriptive view stabilized by further axioms that structured the universe of sets. Replacement and Foundation focused the notion of set, with the first providing the means for transfinite recursion and induction, and the second making possible the application of those means to get results about *all* sets. It

is nowadays almost banal that Foundation is the one axiom unnecessary for the recasting of mathematics in set-theoretic terms, but the axiom is also the salient feature that distinguishes investigations specific to set theory as an autonomous field of mathematics. Indeed it can be fairly said that modern set theory is at base a study couched in well-foundedness, the Cantorian well-ordering doctrines adapted to the Zermelian generative and prescriptive conception of sets. With Replacement and Foundation in place, Zermelo was able to provide natural models of his axioms and to establish algebraic isomorphism, initial segment, and embedding results for his models. Finally Zermelo posited an endless procession of his models, each a set in the next, as natural extensions of their cumulative hierarchies.

Zermelo found a simple set-theoretic condition, being an inaccessible cardinal, that characterizes the ordinal heights of his models, that is those ordinals  $\rho$  such that the predecessors of  $\rho$  are exactly the ordinals of a model.

*Definitions.* An infinite cardinal  $\kappa$  is *singular* if there is an  $x \subseteq \kappa$  of smaller cardinality than  $\kappa$  which is cofinal in  $\kappa$ , that is to say for any  $\alpha < \kappa$  there is a  $\beta \in x$  with  $\alpha \leq \beta$ . An infinite cardinal which is not singular is *regular*

An infinite cardinal  $\kappa$  is a *strong limit* if for any cardinal  $\beta < \kappa$ ,  $2^\beta < \kappa$ .

An infinite cardinal  $\kappa$  is *inaccessible* if it is both regular and a strong limit.

$\aleph_0$  is regular;  $\aleph_1, \aleph_2, \dots$  and generally, all successor cardinals are regular. The limit cardinal  $\aleph_\omega$  is singular, since it has a countable cofinal subset  $\{\aleph_0, \aleph_1, \aleph_2, \dots\}$ . Hausdorff in 1908 had initially entertained the possibility of having a regular limit cardinal. Inaccessible cardinals had later been considered to be a stronger version that arithmetically incorporated power sets, but Zermelo provided the first structural rationale for them, as the delimiters of his natural models.

Inaccessible cardinals are the modest beginnings of the theory of *large cardinals*, a mainstream of modern set theory devoted to the investigation of strong hypotheses and consistency strength. Large cardinal hypotheses posit structure in the higher reaches of the cumulative hierarchy, most often by positing cardinals that prescribe their own inaccessible transcendence over smaller cardinals, and were seen by the 1970s to form a natural hierarchy of stronger and stronger propositions transcending ZFC.

The journal volume containing Zermelo’s 1930 publication also contained Stanisław Ulam’s seminal paper on *measurable cardinals*, which became the most pivotal

of all large cardinals. For a set  $s$ ,  $U$  is a (non-principal) *ultrafilter over  $s$*  if  $U$  is a collection of subsets of  $s$  containing no singletons; if  $x \in U$  and  $x \subseteq y \subseteq s$ , then  $y \in U$ ; if  $x \in U$  and  $y \in U$ , then  $x \cap y \in U$ ; and for any  $x \subseteq s$ , either  $x \in U$  or  $s - x \in U$ . For a cardinal  $\lambda$ , an ultrafilter  $U$  is  $\lambda$ -complete if for any  $D \subseteq U$  of cardinality less than  $\lambda$ ,  $\bigcap D \in U$ . Finally an uncountable cardinal  $\kappa$  is *measurable* if there is a  $\kappa$ -complete ultrafilter over  $\kappa$ . Thus, a measurable cardinal is a cardinal whose power set is structured with a two-valued “measure” having a strong closure property. Measurability embodied the first large cardinal confluence of Cantor’s two legacies, the investigation of definable sets of reals and the extension of number into the transfinite: The concept was distilled from measure-theoretic considerations related to Lebesgue’s measure for sets of real numbers, and it also entailed inaccessibility in the transfinite.

### FORMALIZATION AND MODEL-THEORETIC METHODS

Zermelo’s 1930 publication was in part a response to Skolem’s 1922 advocacy of the idea of framing Zermelo’s 1908 axioms in first-order logic. First-order logic investigates the logic of formal languages consisting of formulas built up from specified function and predicate symbols using logical connectives and first-order quantifiers  $\forall$  and  $\exists$ , these interpreted as ranging over the *elements* of a domain of discourse. (Second-order logic has quantifiers interpreted as ranging over properties, or collections of elements.) First-order logic had emerged in the 1917 lectures of Hilbert as a delimited system of logic potentially amenable to mathematical analysis. Entering from a different, algebraic tradition Skolem had established a seminal result for “metamathematical” methods with the Löwenheim-Skolem Theorem: *If a countable collection of first-order sentences has a model then it has a countable model.*

For set theory Skolem proposed formalizing Zermelo’s axioms in the first-order language with  $\in$  and  $=$  as binary predicate symbols. Zermelo’s *definite* properties were to be those expressible in this first-order language in terms of given sets, and the Axiom of Separation was to become a schema of axioms, one for each first-order formula that has variables allowing for set parameters. As a palliative for taking set theory as a foundation for mathematics, Skolem then pointed out what has come to be called the *Skolem Paradox*: Zermelo’s 1908 axioms cast in first-order logic is a countable collection of sentences, and so if they have a model at all, they have a countable model. (Analogous remarks apply to the latterly adjoined

Axiom of Replacement becoming a schema.) Thus we have the paradoxical existence of countable models for Zermelo’s axioms although they entail the existence of uncountable sets. Zermelo found this antithetical and repugnant. However stronger currents were at work, leading to a further, subtler transmutation of the notion of set mediated by first-order logic and incorporating its relativism of set-theoretic concepts.

Gödel virtually completed the mathematization of logic by submerging metamathematical methods into mathematics. The main vehicle was the direct coding, “the arithmetization of syntax,” in his celebrated 1931 Incompleteness Theorem, which worked dialectically against a program of Hilbert’s for establishing the consistency of mathematics. But starting an undercurrent, the earlier 1930 Completeness Theorem for first-order logic clarified the distinction between the formal syntax and semantics (interpretations) of first-order logic, and secured its key instrumental property with the Compactness Theorem: *If a collection of first-order sentences is such that every finite subcollection has a model, then the whole collection has a model.*

Gödel’s work showed that the notion of the consistency of a mathematical theory has a formal counterpart expressible in the first-order language with function symbols for addition and multiplication. Loosely speaking, a *theory* is a collection of sentences of some first-order language; that a sequence of formulas constitutes a *deduction* can be formalized; and a theory is *consistent* if from it no contradiction can be derived. Gödel’s arithmetization of syntax codes all this into statements about the natural numbers and their arithmetic, yielding a formula

$$\text{Con}(T)$$

asserting the formal consistency of  $T$ , at least for those theories whose sentences can be schematically defined. Gödel famously established through his Incompleteness Theorem that for consistent theories subsuming the arithmetic of the natural numbers,  $\text{Con}(T)$  itself cannot be deduced from  $T$ . However, one may be able to deduce relative notions:

*Definitions.* A sentence  $\sigma$  is *relatively consistent* with a theory  $T$  if  $\text{Con}(T)$  implies  $\text{Con}(T + \sigma)$ .

A sentence  $\sigma$  is *independent* of a theory  $T$  if both  $\sigma$  and its negation are relatively consistent with  $T$ .

Two sentences  $\sigma_1$  and  $\sigma_2$  are *equi-consistent* over a theory  $T$  if  $\text{Con}(T + \sigma_1)$  is equivalent to  $\text{Con}(T + \sigma_2)$ .

These assertions would be established over a weak base theory. For example, in the parlance, that a set-theoretic statement  $\sigma$  is relatively consistent with set theory generally means that  $\text{Con}(\text{ZFC})$  implies  $\text{Con}(\text{ZFC} + \sigma)$ , this itself deducible in (some weak version of) ZFC. *Consistency strength* in set theory can be discussed in these terms, typically for strong set theoretic statements not provable from ZFC: For two set-theoretic statements  $\sigma_1$  and  $\sigma_2$ , the consistency strength of  $\sigma_1$  is least that of  $\sigma_2$  if  $\text{Con}(\text{ZFC} + \sigma_1)$  implies  $\text{Con}(\text{ZFC} + \sigma_2)$ , and so  $\sigma_1$  and  $\sigma_2$  have equal consistency strength if  $\sigma_1$  and  $\sigma_2$  are equi-consistent over ZFC.

Tarski in the early 1930s completed the mathematization of logic by providing his “definition of truth,” exercising philosophers to a surprising extent ever since. Tarski simply schematized truth as a correspondence between formulas of a formal language and set-theoretic assertions about an interpretation of the language and provided a recursive definition of the *satisfaction* relation, when a formula holds in an interpretation, in set-theoretic terms. This response to a growing need for a mathematical framework became the basis for *model theory*. The eventual effect of Tarski’s mathematical formulation of semantics would be not only to make mathematics out of the informal notion of satisfiability, but also to enrich ongoing mathematics with a systematic method for forming mathematical analogues of several intuitive semantic notions. For coming purposes, the following specifies notation and concepts in connection with Tarski’s definition:

*Definitions.* For a first-order language, an interpretation  $N$  of that language (i.e., a specification of a domain of discourse as well as interpretations of the function and predicate symbols), a formula  $\varphi(v_1, v_2, \dots, v_n)$  of the language with the variables as displayed, and  $a_1, a_2, \dots, a_n$  in the domain of  $N$ ,

$$N \models \varphi[a_1, a_2, \dots, a_n]$$

asserts that the formula  $\varphi$  is satisfied in  $N$  according to Tarski’s recursive definition when  $v_i$  is interpreted as  $a_i$ .

A subset  $y$  of the domain of  $N$  is *first-order definable over  $N$*  if there is a formula  $\psi(v_0, v_1, v_2, \dots, v_n)$  and  $a_1, a_2, \dots, a_n$  in the domain of  $N$  such that

$$y = \{z \mid N \models \psi[z, a_1, \dots, a_n]\}.$$

Set theory was launched on an independent course as a distinctive field of mathematics by Gödel’s formulation of the model  $L$  of “constructible” sets, with which he established the relative consistency of the Axiom of

Choice (AC) and the Generalized Continuum Hypothesis (GCH).  $L$  is a transitive class containing all the ordinals that, with the membership relation restricted to it, satisfies each axiom of ZFC as well as GCH. Through  $L$  Gödel established that  $\text{Con}(\text{ZF})$  implies  $\text{Con}(\text{ZFC} + \text{GCH})$  and thus attended to fundamental issues at the beginnings of set theory. In his first, 1938 announcement Gödel described  $L$  as a hierarchy “which can be obtained by Russell’s ramified hierarchy of types, if extended to include transfinite orders.” Indeed with  $L$  Gödel had refined the cumulative hierarchy of sets to a cumulative hierarchy of *definable* sets which is analogous to the orders of Russell’s *ramified* theory. Gödel’s further innovation was to continue the indexing of the hierarchy through *all* the (von Neumann) ordinals to get a model of set theory. In a 1939 note Gödel presented  $L$  essentially as it is presented today: For any set  $x$  let  $\text{def}(x)$  denote the collection of subsets of  $x$  first-order definable over  $x$  according to the previous definition. Then define:

$$L_0 = \emptyset; L_{\alpha+1} = \text{def}(L_\alpha), L_\delta = \bigcup\{L_\alpha \mid \alpha < \delta\} \\ \text{for limit ordinals } \delta;$$

and the *constructible universe*

$$L = \bigcup\{L_\alpha \mid \alpha \text{ is an ordinal}\}.$$

Gödel brought into set theory a method of construction and argument and thereby affirmed several features of its axiomatic presentation. First Gödel showed that  $\text{def}(x)$  and generally first-order definability over set domains is itself definable in set theory, so that in particular the definition of  $L$  can be effected in set theory via transfinite recursion. This significantly contributed to a lasting ascendancy for first-order logic which beyond its *sufficiency* as a logical framework for mathematics was seen to have considerable *operational efficacy*. Gödel’s construction moreover buttressed the incorporation of Replacement and Foundation into set theory. Replacement was immanent in the arbitrary extent of the ordinals for the indexing of  $L$  and in its formal definition via transfinite recursion. As for Foundation, underlying the construction was the well-foundedness of sets, and significantly, Gödel viewed  $L$  as deriving its contextual sense from the cumulative hierarchy of sets regarded as an extension of the *simple* theory of types. In footnote 12 of his 1939 note he wrote, “In order to give  $A$  [that  $V = L$ ] an intuitive meaning, one has to understand by ‘sets’ all objects obtained by building up the simplified hierarchy of types on an empty set of individuals (including types of arbitrary transfinite orders).” Some have been puzzled about how the cumulative hierarchy picture emerged in set-theoretic practice; although there was Mirimanoff,

von Neumann, and especially Zermelo, the picture came in with Gödel's method, the reasons being both thematic and historical: Gödel's work with  $L$  with its incisive analysis of first-order definability was readily recognized as a signal advance, while Zermelo (1930) with its second-order vagaries remained somewhat obscure. As the construction of  $L$  was gradually digested, the sense that it promoted of a cumulative hierarchy reverberated to become the basic picture of the universe of sets.

In a notable inversion, what has come to be regarded as the *iterative conception*, the conception of sets as being built up through stages of construction as schematized by the cumulative hierarchy, has become a heuristic for motivating the axioms of set theory generally. This has opened the door to a metaphysical appropriation in the following sense: It is as if there is some notion of set that is "there," in terms of which the axioms must find some further justification. But set theory has no particular obligations to mirror some prior notion of set, especially one like the iterative conception, arrived at *a posteriori*. Replacement and Choice for example do not quite "fit" the iterative conception, but if need be, Replacement can be "justified" in terms of achieving algebraic closure of the axioms, a strong motivation in the work of Fraenkel and the later Zermelo, and Choice can be "justified" as a logical principle as Zermelo had maintained.

Gödel's proof of the GCH in  $L$ , like Zermelo's proof of the Well-Ordering Theorem, was synthetic and pivotal for the development of set theory. Gödel actually established that if  $\lambda$  is an infinite cardinal and  $x \in L_\lambda$ , then for any  $y \subseteq x$  in  $L$ ,  $y \in L_\lambda$ . The Power Set Axiom was thus tamed in  $L$  leading to the relative consistency of GCH. Replacement played a crucial role not only by providing for the prior extent of ordinals, but also in allowing this first instance of model-theoretic reflection. Reflection properties, which in one form came to be seen as equivalent to Replacement, assert that various properties holding at one level of the cumulative hierarchy holds at an earlier level, and they have been a leading heuristic for motivating large cardinals. Gödel's proof also made a specific, positive use of the Skolem Paradox argument, as he used what are now known as *Skolem functions* to take a *Skolem hull*. Paradox became method, affirming the operational efficacy of first-order logic. Finally Gödel took for the first time what is now known as the *transitive collapse*. Andrzej Mostowski would later state in general terms the result, which is a generalization to well-founded relations and transitive sets of the Mirimanoff–von Neumann result, that every well-ordered set is order-isomorphic to exactly one ordinal with membership. While that result

was basic to the analysis of well-orderings, the transitive collapse result grew in significance from specific applications and came to epitomize how well-foundedness made possible a coherent theory of models of set theory.

In all these ways Gödel's work promoted a further transmutation of, or at least a new relativism about, the notion of set as mediated by first-order logic. By the 1950s ZFC was generally taken to be a theory formalized in first-order logic. The relativism of set-theoretic concepts was brought to the fore, as well as new possibilities for constructions of models of set theory. Results even about definable sets of real numbers would turn on contingencies of relative consistency. Notably, Gödel himself held a "Platonistic" conception of set theory as descriptive of an objective universe schematized by the cumulative hierarchy; nonetheless, his work laid the groundwork for the development of a range of models and axioms for set theory.

Gödel's work with  $L$  stood as an isolated monument for quite a number of years, World War II no doubt having a negative effect on mathematical progress. On the crest of a new generation Dana Scott established a result in 1961 that would become seminal for the theory of large cardinals. Ultrafilters gained prominence in model theory in the late 1950s because of the emergence of the *ultrapower* and more generally *ultraproduct* construction for building concrete models, when Scott made the crucial move of taking the ultrapower of the universe  $V$  itself by an ultrafilter as provided by a measurable cardinal. Such an ultrafilter provided well-founded ultrapowers, and the full exercise of the transitive collapse now led to an inner model  $M$  and an elementary embedding  $j: V \rightarrow M$ .

*Definitions.*  $M$  is an *inner model* if it is a transitive class containing all the ordinals that, with the membership relation restricted to it, satisfies each axiom of ZF.

A class function  $j: V \rightarrow M$  from the universe  $V$  of sets into an inner model  $M$  is an *elementary embedding* if for any set-theoretic formula  $\varphi(v_1, v_2, \dots, v_n)$  and sets  $a_1, a_2, \dots, a_n$ ,

$$\begin{aligned} V \models \varphi[a_1, a_2, \dots, a_n] &\text{ exactly when} \\ M \models \varphi[j(a_1), j(a_2), \dots, j(a_n)]. \end{aligned}$$

(This suggests the general notion of elementary embedding in model theory; the notion cannot be formalized for  $V$  in ZFC, but sufficient schematic approximations can. Below, elementary embeddings are assumed not to be the identity function.)  $L$  is the paradigmatic inner model. Appealing to its definability Scott established: *If there is a measurable cardinal, then  $V \neq L$ .* Large cardinal

hypotheses thus assumed a new significance through a new proof construction, as a means for maximizing possibilities away from Gödel's delimitative universe. The ultrapower construction provided one direction and H. Jerome Keisler soon provided the other of a new characterization that established a central structural role for measurable cardinals: *There is an elementary embedding  $j: V \rightarrow M$  for some inner model  $M$  exactly when there is a measurable cardinal.* Through model-theoretic methods set theory was brought to the point of entertaining elementary embeddings into well-founded models, soon to be transfigured by a new method for getting well-founded *extensions* of well-founded models.

## FORCING

In 1963 Paul Cohen established the independence of the Axiom of Choice from ZF and the independence of the Continuum Hypothesis from ZFC. That is, complementing Gödel's relative consistency results with  $L$  Cohen established that  $\text{Con}(\text{ZF})$  implies  $\text{Con}(\text{ZF} + \text{the negation of AC})$  and that  $\text{Con}(\text{ZFC})$  implies  $\text{Con}(\text{ZFC} + \text{the negation of CH})$ . These results delimited ZF and ZFC in terms of the two fundamental issues at the beginnings of set theory. But beyond that, Cohen's proofs were soon to flow into method, becoming the inaugural examples of *forcing*, a remarkably general and flexible method for extending models of set theory. Forcing has strong intuitive underpinnings and reinforces the notion of set as given by the first-order ZF axioms with conspicuous uses of Replacement and Foundation. If Gödel's construction of  $L$  had launched set theory as a distinctive field of mathematics, then Cohen's method of forcing began its transformation into a modern, sophisticated one. Cohen's particular achievement lies in devising a concrete procedure for extending well-founded models of set theory in a minimal fashion to well-founded models of set theory with new properties but without altering the ordinals. Set theory had undergone a sea-change, and beyond simply how the subject was enriched, it is difficult to convey the strangeness of it.

Cohen's approach was to start with a model  $M$  of ZF and adjoin a set  $G$ , one that would exhibit some desired new property. He realized that this had to be done in a minimal fashion in order that the resulting structure also model ZF, and so imposed restrictive conditions on both  $M$  and  $G$ . He took  $M$  to be a countable standard model, that is a countable transitive set that together with the membership relation restricted to it is a model of ZF. (The existence of such a model is an avoidable assumption in formal relative consistency proofs via forcing.)

The ordinals of  $M$  would then coincide with the predecessors of some ordinal  $\rho$ , and  $M$  would be the cumulative hierarchy  $M = \bigcup_{\alpha < \rho} (V_\alpha \cap M)$ .

Cohen then established a system of terms to denote members of the new model, finding it convenient to use a ramified language: For each  $x \in M$  let  $\dot{x}$  be a corresponding constant; let  $\dot{G}$  be a new constant; and for each  $\alpha < \rho$  introduce quantifiers  $\forall_\alpha$  and  $\exists_\alpha$ . Then develop a hierarchy of terms as follows:  $M_0 = \{\dot{G}\}$ , and for limit ordinals  $\delta < \rho$ ,  $M_\delta = \bigcup_{\alpha < \delta} M_\alpha$ . At the successor stage, let  $M_{\alpha+1}$  be the collection of terms  $\dot{x}$  for  $x \in V_\alpha \cap M$  and "abstraction" terms corresponding to formulas allowing parameters from  $M_\alpha$  and quantifiers  $\forall_\alpha$  and  $\exists_\alpha$ . It is crucial that this ramified language with abstraction terms is entirely formalizable in  $M$ , through a systematic coding of symbols. Once a set  $G$  is provided from the outside, a model  $M[G] = \bigcup_{\alpha < \rho} M_\alpha[G]$  would be determined by the terms, where each  $\dot{x}$  is to be interpreted by  $x$  for  $x \in M$  and  $\dot{G}$  is to be interpreted by  $G$ , so that:  $M_0[G] = \{G\}$ ; for limit ordinals  $\delta < \rho$ ,  $M_\delta[G] = \bigcup_{\alpha < \delta} M_\alpha[G]$ ; and  $M_{\alpha+1}[G]$  consists of the sets in  $V_\alpha \cap M$  together with sets interpreting the abstraction terms as the corresponding definable subsets of  $M_\alpha[G]$  with  $\forall_\alpha$  and  $\exists_\alpha$  ranging over this domain.

But what properties can be imposed on  $G$  to ensure that  $M[G]$  be a model of ZF? Cohen's key idea was to tie  $G$  closely to  $M$  through a system of sets in  $M$  called *conditions* that would approximate  $G$ . While  $G$  may not be a member of  $M$ ,  $G$  is to be a subset of some  $Y \in M$  (with  $Y = \omega$  a basic case), and these conditions would "force" some assertions about the eventual  $M[G]$  that is, by deciding some of the membership questions whether  $x \in G$  or not for  $x \in Y$ . The assertions are to be just those expressible in the ramified language, and Cohen developed a corresponding *forcing relation*  $p \Vdash \phi$ , " $p$  forces  $\phi$ ", between conditions  $p$  and formulas  $\phi$ , a relation with properties reflecting his approximation idea. For example, if  $p \Vdash \phi$  and  $p \Vdash \psi$ , then  $p \Vdash \phi \& \psi$ . The conditions are ordered according to the constraints they impose on the eventual  $G$ , so that if  $p \Vdash \phi$ , and  $q$  is a stronger condition, then  $q \Vdash \phi$ . Scott made an important suggestion simplifying the definition for negation:  $p \Vdash \neg \phi$  if for no stronger condition  $q$  does  $q \Vdash \phi$ . It was crucial to Cohen's approach that the forcing relation, like the ramified language, be definable in  $M$ .

The final ingredient is that the whole scaffolding is given life by incorporating a certain kind of set  $G$ . Stepping out of  $M$  and making the only use of its countability, Cohen enumerated the formulas of the ramified language in a countable sequence (shades of Skolem's

Paradox!) and required that  $G$  be completely determined by a countable sequence of stronger and stronger conditions  $p_0, p_1, p_2, \dots$  such that for every formula  $\varphi$  of the ramified language exactly one of  $\varphi$  or  $\neg \varphi$  is forced by some  $p_n$ . Such a  $G$  is called a *generic* set. Cohen was able to show that the resulting  $M[G]$  does indeed satisfy the axioms of ZF: Every assertion about  $M[G]$  is already forced by some condition; the forcing relation is definable in  $M$ ; and so the ZF axioms, holding in  $M$ , mostly crucially Power Set and Replacement, can be applied to derive corresponding forcing assertions about ZF axioms holding in  $M[G]$ .

The extent and breadth of the expansion of set theory described henceforth far overshadows all that has been described before, both in terms of the numbers of people involved and the results established. With clear intimations of a new and concrete way of building models, set theorists rushed in and with forcing were soon establishing a cornucopia of relative consistency results, truths in a wider sense, some illuminating classical problems of mathematics. Many different forcings were constructed for adding new real numbers and iterated forcing techniques were quickly broached.

Robert Solovay played a prominent role in the forging of forcing as a general method, and he above all in this period raised the level of sophistication of set theory across its breadth from forcing to large cardinals. Solovay proved a result already in 1964 remarkable for its sophistication: Suppose that  $\kappa$  is an inaccessible cardinal; then in an inner model of a forcing extension,  $\kappa$  becomes  $\aleph_1$ , the least uncountable cardinal, every set of real numbers is Lebesgue measurable, and Dependent Choices (a substantial form of AC for bolstering measure) holds. This model offered important insights into the possibilities of measure and the limits imposed by AC. The inaccessible cardinal was thought for some time to be an artifact of the proof, when in 1979 Saharon Shelah finally complemented Solovay's result by showing that if every set of real numbers is Lebesgue measurable and Dependent Choices holds, then  $\aleph_1$  (in  $V$ ) is inaccessible in the constructible universe  $L$ .

Through the 1970s and into the 1980s the forcing method was honed with sophisticated iterated forcing techniques, techniques that established new, more contextualized relative consistency results in the self-generating mainstreams of set theory, *infinitary combinatorics* and *cardinal invariants of the continuum*. Donald Martin formulated an instrumental "axiom," *Martin's Axiom* (MA), in terms of forcing notions, an axiom that became convenient and focal for relative consistency

results. MA together with the failure of CH is relative consistent with ZFC via forcing, and MA directly implies many combinatorial statements in a way analogous to how CH had, and so relative consistency results can be established by drawing direct consequences from MA. A culmination in this direction was the work of Shelah in the 1980s on *proper forcing*, a wide class of forcing notions. Corresponding to MA in this context is the *Proper Forcing Axiom*, an axiom requiring large cardinals to establish its relative consistency. An important barrier that has resisted many efforts is that starting with a model of CH, many iterated forcing constructions have established the relative consistency of various propositions with the continuum being  $\aleph_2$ , but corresponding relative consistencies with the continuum being at least  $\aleph_3$  are not known. Can this be a limitation of forcing, or a delimitation imposed by ZFC?

## LARGE CARDINALS AND INNER MODELS

A subtle connection quickly emerged, already in the 1960s and into the 1970s, between large cardinals and combinatorial propositions low in the cumulative hierarchy: Forcing showed just how relative the Cantorian notion of cardinality is, since one-to-one correspondence functions could be adjoined to models of set theory easily, often with little disturbance. In particular large cardinals, highly inaccessible from below, were found to satisfy substantial propositions even after they were "collapsed" by forcing to  $\aleph_1$  or  $\aleph_2$ , that is correspondence functions were adjoined to make the cardinal the first or second uncountable cardinals respectively. Conversely such propositions were found to entail large cardinal hypotheses in the clarity of an  $L$ -like inner model, sometimes the very same initial large cardinal hypothesis. Thus, in a subtle synthesis, hypotheses of length concerning the extent of the transfinite were correlated with hypotheses of width concerning the fullness of power sets low in the cumulative hierarchy, sometimes the arguments providing *equi-consistencies*. Solovay's Lebesgue measurability result from inaccessibility when complemented by Shelah's result became an equi-consistency, albeit a sophisticated one bringing together Cantor's two legacies, the investigation of definable sets of reals and the extension of number into the transfinite. Other "weak" large cardinals were formulated, sometimes in response to the need of a large cardinal concept to gauge a set-theoretic proposition via equi-consistency. The complementarity also encompassed "strong" large cardinal hypotheses formu-



lated in terms of elementary embeddings and later, new canonical inner models.

Large cardinal hypotheses stronger than measurability were charted out in the late 1960s, motivated not only by the heuristics of generalization but also by those of *reflection*. The direct reflection heuristic is that various properties attributable to the class of all ordinals, since its extent is uncharacterizable, should be attributable already to some cardinal. This heuristic was already at work in Zermelo's 1930 paper and extends the closure provided by Replacement. The more subtle reflection heuristic is that strong large cardinal hypotheses posit elementary embeddings  $j: V \rightarrow M$ , and the closer the target inner model  $M$  is to  $V$ , the stronger the properties that translate and can be reflected between. The *supercompact* cardinals were thus formulated by Solovay and William Reinhardt as global generalizations of measurable cardinals; stronger than these were the *n-huge* cardinals; and the stronger hypotheses still were formulated. There is an ultimate delimitation in this direction that has framed the possibilities: Kenneth Kunen established in ZFC that there can be no elementary embedding  $j: V \rightarrow V$  of the universe into itself. ZFC rallied at last to force a veritable Götterdämmerung for large cardinals.

The theory of these strong hypotheses was developed particularly to investigate the possibilities for elementary embeddings. But what really intimated their potentialities were new forcing proofs, especially from supercompactness, that established the relative consistency of strong existence assertions low in the cumulative hierarchy, at the very least lending these assertions an initial plausibility. The possibility of new complementarity was then brought about through the development of *inner model theory*, the mostly sophisticated part of the theory of large cardinals.

Gödel's  $L$  was the first inner model, and Ronald Jensen dramatically transformed its investigation in the 1960s by refining the first-order definability and Skolem hull arguments to a "fine structure" analysis, extracting important combinatorial principles and establishing new relative consistencies. Inner models of measurability were soon developed, and their interactions and fine structure investigated, and these models would be paradigmatic for inner models of large cardinals: They exhibited in their crystalline clarity akin to algebraic closure the minimal consequences of the large cardinal hypothesis and the maximal structural regularity. In the 1970s, Jensen and Anthony Dodd developed the *core model* for measurability, and this would be paradigmatic for core models of large cardinals: These were inner models that did not

contain the large cardinal, but exhibited the maximal possibilities "up to" the cardinal. The ascent through the large cardinal hierarchy had begun, the inner and core models providing an abiding sense of structure for large cardinal hypotheses.

The development of core models, while quickly developing a life of its own, was initially triggered by work on the Singular Cardinals Problem. With the advent of forcing it had been quickly seen that ZFC imposed little control on the powers  $2^\kappa$  of *regular* cardinals  $\kappa$ , successor or limit, since it became possible to extend a model of set theory by adjoining arbitrarily many subsets of such  $\kappa$  without adjoining any subsets of smaller cardinals. Thus Cantor's Continuum Problem and its generalization to regular cardinals were informed by a general manifestation of method. What about singular cardinals? Powers of singular cardinals seemed much less flexible with respect to forcing, and the *Singular Cardinals Problem* is the general problem of clarifying the possibilities for the function  $2^\kappa$  for singular cardinals  $\kappa$ . Jensen, who found a seminal 1974 result of Jack Silver on powers of singular cardinals "shocking," was directly inspired by it to establish the *Covering Theorem* for  $L$ , easily the most important result of the 1970s in set theory. Very loosely speaking this theorem asserts that unless a surprisingly simple proximity criterion between  $V$  and  $L$  holds, a large cardinal transcendence over  $L$  ensues. It was efforts to extend this result that led to the core models. Through forcing and inner model analysis, results especially of Moti Gitik of the late 1980s established equi-consistency results for simple assertions about powers of singular cardinals and showed remarkable level-by-level connections with large cardinals that affirmed their central place in the investigation of the transfinite.

The extensive research through the 1970s and 1980s considerably strengthened the view that the emerging hierarchy of large cardinals provides *the* hierarchy of exhaustive principles against which all possible consistency strengths can be gauged, a kind of hierarchical completion of ZFC. First the various hypotheses, though historically contingent, form a *linear* hierarchy, one neatly delimited by Kunen's inconsistency result. Typically for two large cardinal hypotheses, below a cardinal satisfying one there are many cardinals satisfying the other, in a sense prescribed by the first. And second, a variety of strong propositions have been informatively bracketed in consistency strength between two large cardinal hypotheses: the stronger hypothesis implies that there is a forcing extension in which the proposition holds; and if the

proposition holds, there is an inner model satisfying the weaker hypothesis.

One of the great successes for large cardinals has to do with perhaps the most distinctive and intriguing development in modern set theory. Although the *determinacy of games* has roots as far back as a 1913 note of Zermelo, the concept of infinite games only began to be seriously explored in the 1960s when it was realized that it led to “regularity” properties for sets of real numbers like Lebesgue measurability.

With  $\omega$  the set of natural numbers let  ${}^\omega\omega$  denote the set of functions from  $\omega$  to  $\omega$ . For  $A \subseteq {}^\omega\omega$ ,  $G(A)$  denotes the following “infinite two-person game with perfect information”: There are two players, *I* and *II*. *I* initially chooses an  $x(0) \in \omega$ ; then *II* chooses an  $x(1) \in \omega$ ; then *I* chooses an  $x(2) \in \omega$ ; then *II* chooses an  $x(3) \in \omega$ ; and so forth:

$$\begin{array}{llll} I : & x(0) & x(2) & \dots \\ II: & x(1) & x(3) & \dots \end{array}$$

Each choice is a *move* of the game; each player before making each of his moves is privy to the sequence of previous moves (“perfect information”); and the players together specify an  $x \in {}^\omega\omega$ . *I* wins  $G(A)$  if  $x \in A$ , and otherwise *II* wins. A *strategy* is a function from finite sequences of natural numbers to natural numbers that tells a player what move to make given the sequence of previous moves. A *winning strategy* is a strategy such that if a player plays according to it he always wins no matter what his opponent plays.  $A$  is *determined* if either *I* or *II* has a winning strategy in  $G(A)$ . The extent of the determinacy of games was investigated through hierarchies of definable sets of reals, and in 1962 the following sweeping axiom was proposed:

**AXIOM OF DETERMINACY.** Every  $A \subseteq {}^\omega\omega$  is determined.

This axiom actually contradicts the Axiom of Choice, as one can get a counterexample  $A$  by “diagonalizing” through all strategies, and so the axiom was intended to hold at least in some inner model to establish regularity properties for sets of real numbers there. In the late 1960s initial connections were made between the Axiom of Determinacy and large cardinals by Solovay, who showed in ZF that the axiom implies that  $\aleph_1$  is measurable, and by Martin, who showed in ZFC that if there is a measurable cardinal, then the analytic sets, the simplest significant sets of real numbers definable with quantifiers ranging over real numbers, are determined. Investigating further consequences of determinacy, a new

generation of descriptive set theorists soon established an elaborate web of connections in the unabashed pursuit of structure for its own sake. Determinacy hypotheses seemed to settle many questions about definable sets of reals and to provide new modes of argument, leading to an opaque realization of the old Cantorian initiatives concerning sets of real numbers and the transfinite with determinacy replacing well-ordering as the animating principle. By the late 1970s a more or less complete theory for the “projective” sets of real numbers was in place, and with this completion of a main project of descriptive set theory attention began to shift to questions of overall consistency.

The investigation of the Axiom of Determinacy spurred dramatic advances in the theory of large cardinals and affirmed their central role in gauging consistency strength. In the 1970s the strength of the methods made possible by the axiom led to speculation that either the axiom was orthogonal to large cardinals or would subsume them in a substantial way. However, large cardinal hypotheses, first near Kunen’s inconsistency and then around supercompactness, were shown to tame Determinacy. By looking at the workings of a proof, Hugh Woodin in 1984 formulated what is now known as a *Woodin cardinal*. Then Martin and John Steel showed that having more and more Woodin cardinals establishes the determinacy of more and more sets in the “projective hierarchy” of sets, sets of real numbers definable with quantifiers ranging over the real numbers. Finally Woodin established by 1992: *the existence of infinitely many Woodin cardinals is equi-consistent with the Axiom of Determinacy*. Woodin cardinals are weaker than supercompact cardinals, closer to measurable cardinals, and in subsequent developments the inner model theory was advanced to getting inner and core models of Woodin cardinals.

Woodin in the late 1990s built on the wealth of ideas surrounding Woodin cardinals and Determinacy and raising them to a higher level proposed a resolution of the Continuum Problem itself. This resolution features the use of arbitrarily many Woodin cardinals, the assimilation of new principles for sets of sets of real numbers, and an unresolved new conjecture about a new “logic” that would complete the picture. Thus structural ideas involving large cardinal hypotheses may circle back to effect an ultimate resolution of the original problem that stimulated the development of set theory.

What about the consistency of large cardinal hypotheses? As postulations for cardinals of properties of the class of all ordinals, they inherit substantial inaccessi-

bility properties from below, but even for large natural numbers given notationally, the meaning of a number is not conveyed by its dogged approach from below but by its mathematical postulation and the sense given it by proof and method. The inner model theory has fortified large cardinals up to Woodin cardinals by providing them with coherent inner models whose structure incisively exhibit their consistency. As for the hypotheses near Kunen's inconsistency, since that result was based on a combinatorial contingency, it could well be that a like inconsistency for a weaker hypothesis can be established. In any case these near-inconsistency hypotheses are less relevant, the forcing proofs applying them to get initial plausibilities having given way to more refined arguments from weaker hypotheses. Moreover the work of Woodin has shown that there is also quite a lot of structure near the Kunen inconsistency, analogous to the descriptive set theory of real numbers.

Stepping back to gaze at modern set theory, the thrust of mathematical research should deflate various possible metaphysical appropriations with an onrush of new models, hypotheses, and results. Shedding much of its foundational burden, set theory has become an intriguing field of mathematics where formalized versions of truth and consistency have become matters for manipulation as in algebra. As a study couched in well-foundedness ZFC together with the spectrum of large cardinals serves as a court of adjudication, in terms of relative consistency, for mathematical propositions that can be informatively contextualized in set theory by letting their variables range over the set-theoretic universe. Thus set theory is more of an open-ended framework for mathematics rather than an elucidating foundation. It is as a field of mathematics proceeding with its own internal questions and capable of contextualizing over a broad range that set theory has become an intriguing and highly distinctive subject.

**See also** Cantor, Georg; First-Order Logic; Gödel, Kurt; Gödel's Theorem; Hilbert, David; Logical Paradoxes; Logic, History of; Modern Logic; Mathematics, Foundations of; Model Theory; Neumann, John von; Russell, Bertrand Arthur William; Second-Order Logic; Tarski, Alfred; Truth.

### Bibliography

- Bartoszyński, Tomek, and Haim Judah. *Set Theory: On the Structure of the Real Line*. Wellesley, MA: A K Peters, 1995.
- Cantor, Georg. *Contributions to the Founding of the Theory of Transfinite Numbers*. *Mathematische Annalen* 46 (1895), 481–512; and II. *Mathematische Annalen* 49 (1897):

- 312–351; with introduction and notes by Philip E. B. Jourdain. Chicago: Open Court, 1915. Reprinted New York: Dover, 1965.
- Dauben, Joseph W. *Georg Cantor: His Mathematics and Philosophy of the Infinite*. Cambridge, MA: Harvard University Press, 1979.
- Ewald, William, ed. *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*. Oxford: Clarendon Press, 1996.
- Gödel, Kurt. *Collected Works*. Vol. 1. New York: Oxford University Press, 1986.
- Gödel, Kurt. *Collected Works*. Vol. 2. New York: Oxford University Press, 1990.
- Hallett, Michael. *Cantorian Set Theory and Limitation of Size*. Logic Guides #10. Oxford: Clarendon Press, 1984.
- Jech, Thomas. *Set Theory*. Berlin: Springer Verlag, 2002.
- Kanamori, Akihiro. *The Higher Infinite: Large Cardinals in Set Theory from their Beginnings*. 2nd ed. Berlin: Springer Verlag, 2003.
- Kanamori, Akihiro. "Zermelo and Set Theory." *The Bulletin of Symbolic Logic* 10 (2004): 487–553.
- Kunen, Kenneth. *Set Theory: An introduction to Independence Proofs*. Amsterdam: North-Holland, 1980.
- Moore, Gregory H. *Zermelo's Axiom of Choice: Its Origins, Development, and Influence*. New York: Springer-Verlag, 1982.
- Moschovakis, Yiannis N. *Descriptive Set Theory*. Amsterdam: North-Holland, 1980.
- Shelah, Saharon. *Cardinal Arithmetic*. Oxford: Clarendon Press, 1994.
- Shelah, Saharon. *Proper and Improper Forcing*. 2nd ed. Berlin, Springer, 1998.
- Van Heijenoort, Jean, ed. *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*. Cambridge, MA: Harvard University Press, 1967.
- Woodin, W. Hugh. *The Axiom of Determinacy, Forcing Axioms, and the Nonstationary Ideal*. Berlin: Walter de Gruyter, 1999.
- Woodin, W. Hugh. "The Continuum Hypothesis I and II." *Notices of the American Mathematical Society* 48 (2001): 567–576 and 681–690.

**Akihiro Kanamori (2005)**

## SEXISM

The term *sexism* denotes any system of beliefs, attitudes, practices, social norms, or institutional forms that functions to create or perpetuate invidious social distinctions among persons on the basis of their actual or presumed sex. This characterization of sexism reflects a widespread consensus among feminist theorists and queer theorists that the phenomenon cannot be understood simply in terms of the prejudices or ill-intentioned behavior of individuals, but rather must be seen as involving wide-ranging social structures, structures that can affect both the meanings and consequences of the actions of individuals, even if such actions are otherwise benign.

Marilyn Frye (1983) has explained, in just these terms, the inadequacy of a definition of sexism as any act or policy involving an “irrelevant or impertinent marking of the distinction between sexes.” She then bids us to consider an employer who refuses to hire a woman for a managerial position on the stated grounds that his employees would not accept the authority of a woman. We may suppose that the supervisor is right about his employees’ attitudes; thus the woman’s sex is, as things stand, relevant to her ability to do the managerial job. By this definition, then, the employer’s act of discrimination would not count as sexist, yet sexism is surely at work, somehow, in this situation.

The problem, Frye argues, is that the definition fails to take account of the ways in which preexisting social conditions can make sex relevant in situations where it need not be, and should not be. In this case, myriad factors had already conspired to create conditions in which the male employees would not be inclined to trust a woman’s judgment, or would not accept her possessing even limited authority over them. Such factors might include: 1) a history of explicit, *de jure* discrimination against women, limiting their participation in public life and their opportunities to assume authoritative social roles; 2) widespread belief (perhaps as a result of number one above) that women are incapable of carrying out the duties of a manager; 3) gender norms that would make it humiliating for men to submit to the authority of a woman. Against such a background, a woman’s sex becomes relevant to the question of whether she can do the job at hand, no matter what other relevant qualifications she possesses.

Another important point illustrated by this example is that sexism does not require bad intentions. In Frye’s case, we need not assume that the employer shares his workers’ prejudices. He may believe the woman to be otherwise qualified, and regret that circumstances make it unwise for him to hire her. Whether he should be criticized for accommodating the sexist views of his employees is certainly a reasonable question, but it is not the same as the question whether the hiring decision was sexist. The point of recognizing sexism is not to indict and punish individuals, but rather to identify and alter all the factors that contribute to the subordination of women, where it is acknowledged that many of these will involve well-entrenched and perfectly mundane social practices.

Richard Wasserstrom (1977) has made the same point, distinguishing what he calls institutional racism and sexism from overt and covert racism and sexism. In the latter two cases, laws or policies are designed with the

explicit intent of allocating unjustified burdens or unwarranted benefits to individuals on the basis of race or sex. In the overt cases, the categories of race and sex are explicitly mentioned, whereas in the covert cases, the categories are represented by surrogates. Jim Crow laws in the United States and the denial of the franchise to women in the United States prior to 1920 exemplify overt racism and sexism, respectively. The use of grandfather clauses after the U.S. Civil War (1861–1865) to disenfranchise formerly enslaved black men, and workplace safety rules that bar from certain jobs anyone capable of becoming pregnant exemplify covert racism and sexism.

Institutional racism and sexism, however, differ from both covert and overt forms in that there need be no intention on anyone’s part to produce a racist or sexist consequence. Wasserstrom recognizes two sub-forms. The first involves regulations or practices that, while apparently neutral, operate against a social reality already configured by racism or sexism, so that their effects are to maintain or reinforce an unjust social hierarchy. This may be the case with the practice of assigning children to schools on the basis of their neighborhoods—if there is *de facto* segregation in housing, then the assignment policy will lead to or sustain racial segregation of schools, even if no one intended that outcome. The second sub-form of institutional racism or sexism involves practices that embody—at a level not readily accessible to consciousness—racist or sexist concepts and presuppositions. Let us consider in this context the question, raised frequently by critics of feminism, whether it is sexist for a man to offer a seat to a woman on a bus.

It is certainly true that many men sincerely regard gestures of this sort as courteous and respectful, and are offended by the suggestion that the gestures denigrate women. Still, if one probes the larger meaning of these customs, it becomes clear that they are part of a system of conventions that symbolically express and prescribe women’s dependency upon and subordination to men. In the first place, there is no plausible moral or empirical rationale for making sex *per se* the criterion for the appropriateness of such a gesture. General moral considerations dictate that any able-bodied person—regardless of their sex—ought to offer a seat to anyone who is visibly incapacitated or subject to physical stress; a person on crutches, someone struggling with packages, or someone obviously exhausted, as might be the case for a woman in the late stages of pregnancy. But these conditions are not connected to sex. The suggestion that women must be accorded respect in virtue simply of their being female is simply peculiar.

There are obvious rationales for social conventions encoding such a stance toward the elderly: Age correlates with experience and wisdom and younger people are generally indebted in various ways to older people. But there is no estimable quality that correlates with femaleness *per se*. By contrast, the gestures and conventions we are considering make perfect sense in light of the prerequisites of a gender system that makes physical weakness normative for women (the “weaker sex”) and physical strength and control normative for men. Such gestures express the relationship that ought to hold between men and women. In a system of social organization in which men hold, or are assumed to hold, real power over women, conventional acts of faux deference by men to women (such as a man’s rising when a woman enters the room) function not as symbols of genuine respect, but rather as expressions of *noblesse oblige*.

Feminist theorists agree that the kinds of gender roles that exist in patriarchal societies are the raw material of sexism, but disagree about whether the elimination of sexism requires the complete dismantling of gender roles, or only their reform. Some feminist theorists argue that any way of attaching systematic social significance to biological sex will inevitably prove sexist. Theorists of this sort include so-called humanistic feminists, who hold that biological sex is a property accidental to, and thus morally irrelevant to one’s humanity (De Beauvoir 1973, Nussbaum 1999, Antony 1998), and dominance theorists, who hold that gender differences are constructed *ex post facto* to mask or rationalize preestablished power disparities (Haslanger 2000, MacKinnon 1987). All such theorists point out that gender roles function to enforce both sexual dimorphism (the demand that one be clearly identifiable and self-identified as either male or female), and compulsory heterosexism (the requirement that one’s erotic interest be focused exclusively on individuals of the opposite gender).

These restrictive social norms are not only deeply oppressive to transgendered and nonheterosexual people, but distorting and limiting for all members of a human society. Theorists who hold this position generally believe that a great deal of the content of gender roles is socially constructed—that there is no natural necessity linking the components of biological sex (morphology, endocrinology, and genetics) to the features of particular gender roles. But the issue of the naturalness of gender roles is in fact orthogonal to the question whether such roles should be socially enforced. Myopic people are biologically (and probably genetically) different from non-myopic individuals; nonetheless, we assign no social

significance to this difference, and in fact acknowledge a social obligation to mitigate the natural consequences of poor eyesight.

Other feminist theorists hold that there is nothing inherently wrong with the existence of gender roles, and that such roles could, in a different social context, be liberatory and beneficial for all. On this view, sexism is constituted by two things: a) the gratuitous attachment of undesirable qualities, such as physical weakness, to the feminine gender role; and b) the widespread devaluation of central elements of that role, such as emotional sensitivity. According to these theorists, sometimes called *gynocentric* or *difference* feminists, facts about the female role in biological reproduction have inherent social significance, and so there is no serious prospect for eliminating social roles erected on the basis of reproductive difference (Young 1985). Central, then, to a feminine gender role will be the social role of mothering. Because the individuals who have been the predominant occupiers of this role are women, and because women have historically lacked social power, the virtues necessary for the proper performance of this role (e.g., empathy, cooperativeness, imaginativeness, nurturance, and altruism) have been devalued.

To dismantle sexism, these feminine virtues must be recognized as being as important to morality as masculine virtues such as impartiality (Gilligan 1982) and greater social support must be provided those who fulfill such typically feminine roles as tending children, caring for the sick, and managing social relationships (Ruddick 1989). Closely allied with gynocentric feminists are *ecofeminists*, who think that women’s greater involvement with the bodily realities of birth, growth, and even death (in their roles as nurses and caregivers) create for women a more intimate relationship with the natural world than men have (Plumwood 1993). This, in turn, makes women more apt than men to strive for ways of life that are harmonious with nature, with nonhuman animals, and with other human beings. All these theorists agree that war and other forms of violence reflect the sexist devaluation of the feminine, and that a proper appreciation of feminine virtues is essential to producing peace.

This dispute within feminist theory about the nature of sexism has implications for social and legal policy. Gynocentric feminists charge that humanistic feminists are guilty of androcentrism—taking the male as the paradigm of the human. If laws and social institutions take no account of differences between men and women, then women will be forever socially and economically disadvantaged by policies and practices centered on male

needs, abilities, and interests. Humanistic feminists counter that they find the content of masculine gender roles just as objectionable as the content of feminine roles, and equally in need of elimination. The revalued gender roles envisioned by gynocentric feminists reflect a romanticized view of female experience, and threaten to legitimate a host of sexist stereotypes and prescriptions. Laws and policies should be based on parameters rationally related to the issue involved, parameters that will sometimes coincide with sex differences, but will more often not.

Feminist theorists have been increasingly concerned with understanding interactions among sexism and a host of other systems of oppressive social division, including racism, heterosexism, class oppression, ageism (invidious division on the basis of age), and ableism (invidious division on the basis of physical capacities), ethnocentrism, and jingoism. Critical legal theorist Kimberle Crenshaw has introduced the notion of intersectionality to capture the *sui generis* character of multidimensional oppression. Postmodern feminist theorists have appealed to this ever-increasing list of interacting parameters of identity to deconstruct categories such as sex and race, arguing that no one is simply or straightforwardly a woman or a black person, but that the self is essentially fragmented and fluid. However, they acknowledge the difficulty of making sense of oppression without appeal to such categories.

**See also** Affirmative Action; Analytical Feminism; Feminist Legal Theory; Feminist Philosophy of Science: Contemporary Perspectives; Feminist Social and Political Philosophy; Frye, Marilyn; Heterosexism; Racism; Violence.

### **Bibliography**

- Antony, Louise. "Back to Androgyny: What Bathrooms Tell Us about Equality." *Journal of Contemporary Legal Issues* 9 (1998): 1–20.
- Crenshaw, Kimberlé. "Demarginalizing The Intersection of Race and Sex." *University of Chicago Legal Forum* (1989): 139–167.
- De Beauvoir, Simone. *The Second Sex*. Translated by H. M. Parshley. New York: Bantam Books, 1973.
- Frye, Marilyn. "Sexism." In *The Politics of Reality*. Freedom, CA.: The Crossing Press, 1983.
- Gilligan, Carol. *In A Different Voice: Psychological Theory and Women's Development*. Cambridge, MA: Harvard University Press, 1982.
- Haslanger, Sally. "Gender, Race: (What) Are They? (What) Do We Want Them to Be?" *Noûs* 34 (1) (2000): 31–55.

- MacKinnon, Catherine. "Difference and Domination: On Sex Discrimination." In *Feminism Unmodified: Discourses on Life and Law*. Cambridge, MA: Harvard University Press, 1987.
- Nussbaum, Martha. "Women and Cultural Universals." In *Sex and Social Justice*. Oxford: Oxford University Press, 1999.
- Plumwood, Val. *Feminism and The Mastery of Nature*. London: Routledge, 1993.
- Ruddick, Sara. *Maternal Thinking: Toward a Politics of Peace*. New York: Basic Books, 1989.
- Wasserstrom, Richard. "Racism, Sexism, and Preferential Treatment: An Approach to the Topics" *UCLA Law Review* (1977): 581–615.
- Young, Iris. "Humanism, Gynocentrism, and Feminist Politics" *Women's Studies International Quarterly* 3 (1985): 173–183.

*Louise M. Antony (2005)*

## SEXTUS EMPIRICUS

(third century CE)

Sextus Empiricus was almost certainly, as his name suggests, a doctor of the empiricist school, which flourished from the third century BCE until at least the third century CE. His dates are very uncertain, but he probably lived and worked, perhaps in Rome, sometime early in the third century CE. He is mentioned as a prominent skeptic in Diogenes Laertius's *Lives of the Philosophers* (DL) 9.116, written probably in the third century; but the men Diogenes names as his teacher and pupil, Herodotus of Tarsus and Saturninus, are even more obscure. He was certainly known as the authoritative source of skeptical argument a century later, when Saint Gregory of Nazianzus blamed him for the "vile and malignant disease" of arguing both sides of a question that was infecting the church. How original he was is unclear—it is hard to tell partly because our other sources for skepticism are so exiguous—but he seems to have been more of a compiler than an original thinker; and in any case it is to his preservation of a large body of skeptical argumentation, whatever its provenance, that his importance in the history of philosophy is due.

His best-known work is *Outlines of Pyrrhonism* (PH), a digest of the arguments and general strategy of that version of skepticism named for Pyrrho of Elis. The precise relationship between the position outlined by Sextus and that of Pyrrho is unclear—Sextus himself says that his philosophy is called "Pyrrhonism" because "he seems to have applied himself to skepticism more thoroughly and conspicuously than any of his predecessors" (PH 1.7). PH summarizes what Sextus presents at much greater length in another work that also survives, at least in part. This consists of the five surviving books generally

(if misleadingly) known as *Against the Professors* (M 7–11, subdivided as *Against the Logicians* (M 7–8), *Against the Physicists* (M 9–10), and *Against the Ethicists* (M 11). M 7–8 corresponds roughly to an expanded version of PH 2, while M 9–11 is summarized in PH 3. Whatever answered to the general treatment of skepticism in PH 1 is lost. The texts known as M 1–6 form a different treatise, written with a rather different aim, consisting of a series of essays directed against the practices (and practitioners) of six of the seven “liberal arts,” in order, grammar, rhetoric, arithmetic, geometry, astrology, and music. Sextus also tells us that he wrote medical texts as well, but these are lost.

Book 1 of PH presents a general account of skeptical practice. Sextus is careful to avoid any suggestion that what he is presenting is a doctrine, or anything else with uncomfortably Dogmatic overtones. “Dogmatist” was the general name given by skeptics to their opponents who held positive, or even negative, views (the term also functions as an umbrella term to describe theoretically minded, as opposed to empirical, doctors; and was probably borrowed from medical terminology). At the beginning of PH, Sextus presents skeptics as starting out from the same position as all other inquirers: They seek to assuage their disquiet by finding out the truth about things. But in any search there are three possible outcomes: one may (a) claim that one has found what one was looking for; (b) deny that it can be found, saying that it is inapprehensible; or (c) simply keep on searching. Option (a) is the position of the Dogmatists (Stoics, Epicureans, Platonists, and Aristotelians). Option (b) Sextus ascribes to the Academic skeptics, claiming (controversially) that they positively deny that things are apprehensible (in the sense of certainly knowable) as to their actual natures, while also claiming that certain positions, in regard to both factual and evaluative matters are “more plausible” than others (PH 1 236–241); and both of these positions are unacceptably Dogmatic from a Sextan perspective. Moreover, even though both schools report that *epochē* (suspension of judgment) regarding things that are naturally nonevident is the natural conclusion of the inquiry, the Academics present this as a goal, and as a good thing, whereas the Pyrrhonist has no attitude to it at all—it is simply something that happens, although it seems to bring tranquility (the acquisition of which was the initial object of the search) in its train (PH 1.21–30). The Pyrrhonist follows option (c), and keeps on inquiring, following the appearances, but suspending judgment about what, if anything, might lie behind them (PH 1.13–24).

Sextus is acutely aware of the dangers of incoherence involved in this presentation of a life without commitment—he cannot consistently *recommend* it (since that would involve supposing it to be objectively good, or at least choiceworthy); and he cannot claim that, as a matter of fact, following it will have the desired therapeutic effect of removing anxiety. But he can (undogmatically) report his own experiences; it seems to him that this is how things have gone. Moreover, he is moved (so he tells us) by benevolence: Seeing the Dogmatists suffering from their vain pretensions to knowledge, he seeks to cure them (PH 3.280–281), not because he positively affirms that it is good to do so, but simply because he finds himself so moved. In the same vein, skeptics adhere (undogmatically) to “a quadripartite practice of ordinary life,” since “we cannot remain wholly inactive”: they follow “the guidance of nature, the constraint of the affections, the tradition of laws and customs, and the instruction of the arts” (PH 1.23). The skeptic has a “criterion of action”—the appearances—but no “criterion of truth” (PH 2.13–79). Sextus thus shows himself sensitive to the sort of objection made famous by David Hume, but anticipated in the Greek tradition, that skepticism is fatal.

PH 1.31–163 presents a version of the Ten Modes of Aenesidemus, but in a manner that betrays the later influence of Agrippa. The Modes are collections of considerations designed to (or rather, which have been found to) induce *epochē* on all nonphenomenal matters (the causal language is important: there can be no inference as such for the Pyrrhonist). They consist in the collection of “oppositions”—Sextus describes skepticism as “a capacity for oppositions”: PH 1.8—cases where (apparently)  $x$  appears  $F$  to  $y$  (or in circumstances  $C$ ), but not- $F$  to  $z$  (or in circumstances  $C^*$ ); since there can be no non-question-begging way of deciding on the superiority of any one of the opposed appearances over any other (that is, we have no criterion) “we are moved to suspend judgment” (PH 1.78; cf. 1.89, 99, 117). The skeptic will adduce considerations on both sides of any question to promote “equipoise of argument” (PH 1.8, 190), not in order to support or undermine one side or the other. If you claim  $p$ , skeptics will adduce reasons why not- $p$ , not because they believe them *in propria persona*, but simply because their benevolence compels them to. All skeptical argument is dialectical. Thus when Sextus produces arguments against proof (PH 2.134–192), he does so not because he *believes*, inconsistently, in the capacity of proof, as the Dogmatists allege (PH 2.185–186); rather it is the Dogmatists who, insofar as they believe the canons of rational argument, must find those canons undermined from within. The skeptic has no beliefs about them

at all (PH 2.187–192). And it is in this manner that Sextus deploys the vast bulk of his argumentation in PH 2–3 and M 7–11, for and against particular philosophical and scientific positions.

In a similar vein, Sextus discusses the “skeptical slogans,” such as “no more [so than not so],” *ouden mallon* (PH 1.188–191). The phrase *ouden mallon* had been used by earlier philosophers (including Democritus, Plato, Protagoras, and Pyrrho), but to signal, non-skeptically, that some things really *were* no more *F* than not-*F*. For the skeptic, it functions simply as a marker of a refusal to say, one way or another. Equally, when skeptics say “I determine nothing,” or “all things are undetermined” (PH 1.197–199), they do not assert that nothing is determinable; these are merely expressions of how things seem. Indeed, the skeptical slogans apply to themselves: skeptics determine nothing, not even that they determine nothing (PH 1.206–209); here again they can avoid the charge of “negative Dogmatism” (option (b) above) they level at the Academics.

At PH 1.210–241, Sextus seeks to distinguish Pyrrhonism from other superficially similar philosophies by stressing the fact that all of them slide into Dogmatism. Thus Pyrrhonism is not relativism, at least if that positively affirms that everything is relative; the Pyrrhonist appeals to the relativity of appearances, but draws no ontological conclusion therefrom. Curiously, Sextus even distinguishes his practice from that of the empiricist doctors. The latter follow the appearances; and make use of the type of sign (“commemorative”) that Sextus allows (PH 2.100–102), in which something evident is a sign of something else that is only temporarily nonevident, that is, whose existence can be confirmed by further investigation, as when smoke is a sign of (concealed) fire. (Sextus

rejects “indicative signs,” whereby Dogmatists seek to infer to the hidden internal structures of things on the basis of evident phenomena, on the grounds that there can be no noncontroversial inference of such a kind: PH 2.104–133; M 8 199–300). But they also developed a complex epistemology of reasonable expectation, based upon personal observation, reportage, and argument form analogy. And for Sextus, this strays too far toward a Dogmatic supposition that certain outcomes really are more likely than others. For this reason, he prefers the methodic school of medicine as a model, since this school also makes no affirmations, and simply follows the “quadripartite practice of ordinary life” (PH 236–141). In this vein, in M 1–6, Sextus allows that it is fine to practice some skill, as long as one does so undogmatically, that is without commitment to any supposed deep truths that the skill relies upon. This is the sense in which the skeptic may follow “the instruction of the arts,” and how Sextus may consistently be a (type of) doctor.

*See also* Aenesidemus; Agrippa; Ancient Skepticism; Pyrrho.

## ***Bibliography***

### PRIMARY WORKS

Sextus Empiricus. *Outlines of Skepticism*. Translated by Julia Annas and Jonathan Barnes. Cambridge, U.K.: Cambridge University Press, 1994.

### SECONDARY WORKS

Bailey, Alan. *Sextus Empiricus and Pyrrhonian Scepticism*.

Oxford: Oxford University Press, 2002.

Barnes, Jonathan. *The Toils of Scepticism*. Cambridge, U.K.:

Cambridge University Press, 1990.

Hankinson, R. J. *The Sceptics*. London: Routledge, 1995.

***R. J. Hankinson (2005)***