PROOF, LOGIC AND FORMALIZATION

EDITED BY
MICHAEL DETLEFSEN

ROUTLEDGE

PROOF, LOGIC AND FORMALIZATION

This book addresses various problems associated with finding a philosophically satisfying account of mathematical proof. It brings together many of the most notable figures currently writing in the philosophy of mathematics in an attempt to explain why it is that mathematical proof is given prominence over other forms of mathematical justification. The difficulties that arise in accounts of proof range from the rightful role of logical inference and formalization to questions concerning the place of experience in proof and the possibility of eliminating impredicative reasoning from proof.

The editor, Michael Detlefsen, has brought together an outstanding collection of essays, only two of which have previously appeared. Essential for philosophers and historians of mathematics, it is also recommended for philosophically inclined logicians and philosophers interested in the nature of reasoning and justification. A companion volume entitled *Proof and Knowledge in Mathematics,* edited by Michael Detlefsen, is also available from Routledge.

PROOF, LOGIC AND FORMALIZATION

PROOF, LOGIC AND FORMALIZATION

Edited by

Michael Detlefsen

London and New York

In memory of Kenneth Lee Steel, Born 18 March 1947, Lincoln, Nebraska, Died 28 February 1967, Bong Son, South Viet Nam.

"Therefore I have uttered what I did not understand, things too wonderful for me, which I did not know. 'Hear, and I will speak; I will question you, and you declare to me.'"

Job 42

CONTENTS

NOTES ON CONTRIBUTORS

David Auerbach teaches philosophy and logic at North Carolina State University in Raleigh.

John P.Burgess received his doctorate from the Group in Logic and the Methodology of the Deductive Sciences at the University of California at Berkeley, and is now Professor of Philosophy at Princeton University. He is an editor of the *Journal of Symbolic Logic* and the *Notre Dame Journal of Formal Logic,* and a frequent contributor to journals and anthologies in mathematical and philosophical logic and the philosophy of mathematics.

Michael Detlefsen is Professor of Philosophy at the University of Notre Dame and Editor-in-Chief of the *Notre Dame Journal of Formal Logic.* His areas of special interest are the philosophy of logic and mathematics. He is the author of *Hubert's Program: An Essay on Mathematical Instrumentalism* (Reidel, 1986). Currently, he is working on a book on intuitionism, and coauthoring another with Michael Byrd on Gödel's theorems.

Glen Helman, formerly a member of the Philosophy Department of the Pennsylvania State University, is now Assistant Professor of Philosophy at Wabash College. His publications concern the equivalence of proofs and their relation to the extended lambda calculus.

Daniel Isaacson is currently University Lecturer in the Philosophy of Mathematics at Oxford University and Fellow of Wolfson College, Oxford. He has been lecturer in philosophy at the University of Washington in Seattle and visiting professor of philosophy at the University of California, Berkeley, and has held research posts at the Rockefeller University in New York and St John's College, Oxford. He was an undergraduate at Harvard, and took his doctorate at Oxford.

Charles Parsons is Professor of Philosophy at Harvard University. From 1965 to 1989 he was on the faculty of Columbia University. He is the author of *Mathematics in Philosophy* (Cornell University Press, 1983) and of papers on proof theory, intensional logic, philosophy of logic and mathematics and Kant.

Richard Tieszen received his Ph.D. in philosophy at Columbia University and is presently Assistant Professor at San Jose State University, located in the Silicon Valley. His research interests include logic and logic programming, philosophy of mathematics, and cognitive science.

Robert S.Tragesser is currently Assistant Professor of Philosophy at Barnard College in Columbia University, and Research Associate at the CUNY Graduate Center. He is the author of two books, *Phenomenology and Logic* (Cornell University Press, 1977) and *Husserl and Realism in Logic and Mathematics* (Cambridge University Press, 1984). He is at work on a book, *Logical Imagination,* which will contain (among other things) an elaboration of the ideas presented in his paper in this volume.

PREFACE

Among the features of mathematical thought that most invite the philosopher's attention is its regimentation of justificatory procedures; its canonization of proof as the preferred form of justification. This regimentation is made the more striking by the fact that it is clearly possible to have other kinds of justification (even some of very great strength) for mathematical propositions; a fact which at least suggests that concern for truth and/or certainty alone cannot account for the prominence given to proof within mathematics. This being so, arriving at some understanding of the nature and role of proof becomes one of the primary challenges facing the philosophy of mathematics. It is this challenge which forms the motivating concern of this volume, and to which it is intended to constitute a partial response (on a variety of different fronts).

One possibility, of course, is that it signifies nothing distinctive in the epistemic aims of mathematics at all, but is rather the outgrowth of certain purely social or historical forces. However, since regimentation of justificatory practice is the kind of thing that is at all likely to arise only out of consideration for epistemic ideals, to attribute it to nonepistemic, socio-historical factors would require special argumentation, and there is currently little if any reason to suppose such argumentation might be given. Thus, accounts of the ascendancy of proof in mathematics that make appeal to some epistemic ideal(s) would appear to be more promising.

One consideration which seems to have weighed heavily in favor of proof as the preferred form of justification is the traditional concern of mathematics for rigor. This drive, in turn, seems to have derived from one of two (in some ways related) impeti. The first of these—an old and persistent one, with a bloodline reaching from Aristotle through Leibniz and on down to Frege —s based on the idea that the aim of a justification is not merely, nor even primarily, to establish the certainty of its conclusion, but rather to identify the metaphysical grounds or origins. This induces a hierarchy of truths at the base of which are located the ur-truths or *axioms.* The epistemic ideal is to be able to trace a given truth all the way back to the ur-truths which found it, thus obtaining a complete account of why it is true. Rigor is essential to this enterprise because, without it, one cannot get a clear distillation of those truths that are being used (and perhaps even *need* to be used) to ground a given truth.

A historically more influential idea—also associated with Aristotle (in particular, with his conception of a deductive or demonstrative science), as well as with Euclid and, in the modern period, with Descartes—has a less metaphysical, more purely epistemic foundational thrust. It too searches for ultimate truths, but not in the sense of final metaphysical grounds of truth. Rather, it seeks to find premises that are epistemically ultimate—truths, that is, that are (at least in certain respects) epistemically unsurpassed. In this type of scheme, the axioms are not taken to be metaphysically responsible for the truth of the theorems, though there is still a sense in which they may be said to show why they (i.e. the theorems) are true.

More recently, these traditional conceptions have given way to less foundationally oriented conceptions of mathematical justification. Axioms are no longer taken as giving the metaphysical grounds for the theorems. Nor are they taken to be insurpassably, or even insurpassedly, evident. Rather, they are seen only as being evident enough (i) to make a search for further evidence seem unnecessary, or perhaps (ii) to remove sufficiently much of whatever rational doubt or indecision it seems possible and/or desirable to remove from the theorems, or maybe just (iii) to be attractively simple and economical, and evident enough to serve the purposes at hand, etc.

This shift away from foundationalist conceptions of justification has not, however, brought with it an associated shift away from rigor. Indeed, with the vigorous development of techniques of *formalization* that has taken place in this century, demands for rigor have increased to a point where it is now the reigning orthodoxy to require that, to be genuine, a proof must be formalizable. This emphasis on formalization is based on the belief that the only kinds of inferences ultimately to be admitted into mathematical reasoning are *logical* inferences (i.e. inferences which themselves make appeal to no mathematical subject-matter), and the conviction that the only way ultimately to be assured that one's reasoning does not make use of unrecognized assumptions is to divest it of all appeals to meaning—even the meaning of logical terms.

These emphases on the role of logic in mathematical reasoning and the ability to arrive at a verifiable description of mathematical reasoning that does not make use of appeals to meaning, together with the question of how best to account for their epistemological purposes, are matters of central concern to the essays in this volume. The rightful role of logical

inference in proof is a common theme of the papers by Burgess, Helman, Tieszen, and Tragesser; the latter suggesting that logic has little if any role to play in proof, the others suggesting what end up to be one form or another of the contrary view.

Burgess and Helman argue for views which see classical logic as occupying a place of central importance in mathematical reasoning. For Burgess this is a result of the fact that classical logic gives that account of the inferential structure of mathematical reasoning which fares best with respect to two desiderata which he takes to form the prime determinants of any satisfactory theory of math-ematical reasoning; namely, (i) whether it gives an empirically successful (if also idealized) *description* of the natural epistemic behavior of mathematicians, and (ii) whether it uses a *method* of description which itself is amenable to precise mathematical investigation and elaboration. These desiderata are emphasized because Burgess takes them to be essential to truly *scientific* account of proof, and he takes that to be the kind of theory of proof that we ought to be most intent upon producing.

In Helman's argument, the importance of logic to proof is derived from a structural similarity which he sees as existing between systems of abstract proofs (modulated by an equivalence relation defined on them), on the one hand, and systems of information employed in assessing the justification of beliefs, on the other. This structural similarity arises from the fact that a justification for a conclusion is identified by reference to justifications for its ultimate premises. Helman suggests that the structure thus induced between premise-justifications and conclusion-justifi- cations may correspond to a certain kind of algebraic structure exhibited by systems of abstract proofs, which algebraic structure then gives rise to the rules of logical inference. If this is right, then a description of this justificational structure would serve as a means of *interpreting* the corresponding algebraic structure, and so would provide an epistemological basis for logic.

Tieszen's paper implies a brief for a kind of intuitionist logic— though this arises only to the extent that it is called for by his more basic commitment to a Husserlian epistemology for proof, which conceives of it as fundamentally a case of *intention-fulfillment.* Since intentions are directed toward prepositional contents, and since such contents are quite often "composite" in character, it is natural to look for a logic of intention-fulfillment that is induced by the component-composite relationships that exist between the contents of simpler and more complex intentions. It is this sort of logic that Tieszen sketches, and so, like Helman, he can be seen as looking for a logic based on certain basic structural features of (mathematical) justification.

Issues of formalization form another point at which the interests of several of the papers published here converge. Regarding the question of the formalizability of mathematical reasoning, the papers by Tieszen and Tragesser present negative responses (though only when mathematical proof is considered in its justificatory capacity and not as, say, the object of some metamathematical inquiry such as that of proving consistency). Burgess and Helman suggest the opposite view.

For Tieszen, epistemically genuine proof is unformalizable because it is essentially an *experience*—specifically, an experience of the fulfillment of an intention. For Tragesser, genuine proof is not formalizable by the usual means because such formalized proof tends to mask, or at least not properly to bring out, the element of *understanding* that is crucial to it and which renders the logical inferences that typify formalized proof epistemically otiose.

Taking the opposite view, Helman, as previously noted, argues that there is a profound connection between the epistemic structure of a proof and that logical structure which comes to be highlighted through its formalization. Thus, a proofs lacking the sort of logical structure revealed in its formalization(s) is tantamount to its lacking the sort of structure that Helman suggests is a deep and deeply important part of justificatory reasoning in mathematics.

Burgess, too, suggests an opposing view since, on his view, a proof which is not (at least in a suitably broad sense) formalizable is not accommodatable to the methodological constraint governing the scientific investigation of proof; namely, that it be "mathematizable" (i.e. open to investigation by a method possessing the kind of precision that typifies modern science). (It is perhaps worth noting in passing, however, that such a defense for formalization does not in and of itself make a case for the "effectively decidable" character of the common formalization of such basic metamathematical relations as "formula *x* is derivable from formulae *y* in system *z*." Mathematically precise but not effectively decidable versions of such notions are well known from the literature.)

Different concerns regarding formalization form the foci of the three papers on Gödel's theorems that appear in this volume (namely, those by Isaacson, Auerbach and myself). The concern of Isaacson's paper is whether PA is an adequate (i.e. sound and complete) formalization of a certain kind of truth possessable by sentences formulable in the language of PA. As is well known, of course, Gödel's First Theorem shows that PA cannot be a complete formalization of the *classical* truths formulable in its language. The premise of Isaacson's paper, however, is that this does not rule out the possibility that there are *other* interesting notions of truth with respect to which PA *is* complete. Indeed, Isaacson claims to have found one in a certain *epistemic* notion of truth which he calls "arithmetic truth."

Isaacson produces *inductive* evidence for the claim that this notion of truth is formalized by PA by considering the effects of adding various   -rules to the latter. Specifically, he argues that (i) those   -rules which constitute forms of inference sanctioned by his notion of truth add nothing to PA, and (ii) those   -rules that do add something to PA are not sanctioned by this notion of truth. He concludes, therefore, that what look to have been possible means of upsetting the fit between PA and

"arithmetic truth" (namely, the various -rules examined) in fact are not, and that this gives us (further) inductive support for regarding the former as a complete formalization of the latter.

In his paper "How to say things with formalisms," David Auerbach's general concern is that of determining what is required for adequate *representation of* formal notions of proof. More specifically, since the conditions for adequate representation, being dependent upon the purposes for which a representation is wanted, cannot be expected to be invariant, he is concerned to defend those conditions (the so-called "Derivability Conditions") that are typically used to prove generalized versions of Gödel's Second Theorem. His defense, in effect, is to say that these conditions are satisfied by any representation of a *usual* definition of formal proof that preserves its (i.e. the definition's) *structure* or *form.* He then goes on to argue that preservation of this definitional form is a good thing, and that therefore Gödel's Second Theorem is a stable accompaniment of good representation of the basic metamathematical notions.

My paper considers a different kind of motive for violating the Derivability Conditions; namely, that which is promoted *not* by modifications of what is to count as an *adequate representation of* a notion of formal proof, but rather by modifications of the very concept of theory (and proof) itself. The particular study here is set in the context of a more general discussion of Hilbert's Program; specifically, one centered on the question of whether an argument against Hilbert's Program might better be based on Gödel's *first* rather than his *second* theorem—an idea that has been suggested by several people writing on the subject. The basis for this suggestion, at least in part, appears to be due to a difference in the stability of the two theorems which may be described roughly as follows: Gödel's Second Theorem is sensitive to the kind of modification alluded to two sentences ago while Gödel's First Theorem is not. Hence, the latter would appear to be the more stable of the two phenomena, and thus the preferable basis for an argument against Hilbert's Program.

In opposition to this, I argue that the conservation condition ruled out by the first theorem is an unreasonable constraint to place on Hilbert's systems of ideal mathematics. This, of course, leaves the second theorem to contend with and so, continuing a line of argument initiated in my earlier book on the subject, I go on to indicate certain gaps in the alleged refutation of Hilbert's Program from the second theorem, in particular those resulting from failure adequately to take account of the possibilities of merging Hilbert's instrumentalistic conception of ideal proof with various nonmonotonic, "consistency-minded" conceptions of proof.

Matters of a somewhat broader epistemological nature arise in the papers of Parsons and Tragesser. Parsons considers a question taken up earlier by Poincaré. It concerns the character of impredicative reasoning and its overall place in the production of justifi cation in mathematics. Of particular concern is whether impredicative reasoning is, by its very nature, viciously circular, as Poincaré alleged. At first sight, one might take this to be a somewhat peripheral concern in the philosophy of mathematics, but Parsons produces an argument with quite the contrary effect. On his view, impredicative reasoning turns out to be surprisingly pervasive. Even the most basic arithmetical concept formation, he argues, is impredicative in certain respects. And from this he draws the conclusion that unless we are prepared to adopt a very hard line regarding the possibility of mathematical knowledge (in particular, one that would rule out much of elementary arithmetic), we cannot accept Poincaré's unfavorable view of the epistemological character of impredicative reasoning. Thus, on Parsons' view, skepticism regarding impredicative reasoning brings with it a much more general skepticism concerning mathematical knowledge.

Tragesser takes up the old question concerning the character of the warrant produced by mathematical proof. He argues that (i) well-described *possible* proofs have the same epistemic force as *actual* proofs, and that (ii) this fact gives rise to a fundamental distinction between empirical and mathematical warrant since, on his view, a well-described possible empirical warrant does *not* have the same epistemic force as an actual empirical warrant. He then uses this argument to provide a basis for asserting the necessity of mathematical theorems on the grounds that, if a well-described possible proof of **p** suffices for its actual proof, then there can be no possible proof of the denial of a proven proposition **p**—a state of affairs from which he believes the necessity of **p** follows.

With the exception of the papers by Parsons and myself, none of the essays appearing here has been published before, and Parsons' is a substantial revision of its forerunner. This represents a lot of hard work done specifically for this volume, and I should like to express my sincerest thanks to the contributors for all their efforts. My thanks also to Michael Byron, who corrected the proofs and compiled the index for this collection. They have made my job as editor both a pleasure and a privilege, and I am grateful to them all.

<div align="right">

Michael Detlefsen
Notre Dame, Indiana
December 1990

</div>

1
PROOFS ABOUT PROOFS A DEFENSE OF CLASSICAL LOGIC
Part I: the aims of classical logic

John P.Burgess

Let us remember that *we are mathematicians...*

<div align="right">Hilbert</div>

SUMMARY

The aims and claims of classical logic are expounded, in preparation for its defense against the counterclaims of non-classical logic.

INTRODUCTION

The question, "Which is the right logic?" invites the immediate counter-question, "Right for what?" Hence a defense of classical logic must consist of two parts: (i) a preparatory exposition of the general aims of classical logic, and (ii) a defense against specific counterclaims of nonclassical logics. The present paper contains the first part of such a defense, expounding and defending a conception of the aims of classical logic.

According to this conception, the object and the method of investigation of classical logic are both mathematical proof. Given its mathematical method, it must be descriptive rather than prescriptive, and its description must be idealized. The appropriateness of its idealization is to be evaluated by the success of its application in the investigation of its object, the scope and limits of mathematics.

The conception just indicated, and expounded below, will be defended as being both historically and scientifically important: historically, because it approximates the conception of the founders of classical logic; scientifically, because it attributes to classical logic an object of investigation of considerable importance, and because its method of investigation of that object has had considerable success. Attacks on classical logic based on the attribution to it of some different aim are misconceived, according to the line of defense to be erected in the present paper.

Detailed consideration of specific logics usually conceived of as alternatives to the classical, such as partial, many-valued, relevance/relevant, intuitionistic, free, and quantum logics, will be deferred to a projected sequel.

I

Classical logic is to be conceived of as having mathematical proof as its object of investigation. Specifically, its object is classical, as opposed to constructivist, mathematics; and deduction as distinguished from axiomatics, how proofs proceed as distinguished from where proofs begin.

Traditional logic was a theory of deduction: in part, as in the *Sophistic Refutations,* deduction as practiced in philosophical debate; and in part, as in the *Prior Analytics,* deduction as practiced in mathematical proof. Classical logic was developed by Frege, Peano, Russell, Hilbert, Skolem, Gödel, Tarski, and other founders as an extension of traditional logic mainly, if not solely, about proof procedures in mathematics. This point is so clear from the contents, and indeed the very titles, of their major works that it perhaps need not be argued here.

Nor ought it to be needful to argue the scientific importance of deduction in classical mathematics as an object of investigation. It ought not to be necessary to argue the importance of rigorous deduction in mathematics, or of rigorous deductive mathematics in science. Yet in the present climate of opinion, what ought not to be necessary perhaps is. For much of the recent philosophical literature on mathematics and science, by authors otherwise as diverse as Lakatos (1976), Kitcher (1984), and Tymoczko (1987), has, to put it positively, emphasized factors other than rigorous deduction. As this trend in the recent literature has been partly a reaction against and corrective to an overemphasis on rigorous deduction in earlier literature, any defense of the continued scientific importance of the object of investigation of classical logic must begin by acknowledging that other factors are also important in the investigation of mathematics and the role it plays in science.

First, it must be acknowledged that the requirements of rigor pertain to the context of justification, publication for collective evaluation by a community of colleagues, and not to the context of discovery, private mental processes of individual researchers. No one discovers a theorem by first discovering the first step of the proof, second discovering the second step of the proof, and so on. The role of inductive, analogical, heuristic, intuitive, and even unconscious, thought in the context of discovery has been emphasized by all mathematicians discussing mathematics, and notably in the books of Hadamard (1954) and Polya (1954). Much the same point is at issue in the dicta of Poincaré, "The sole function of logic is to sanction the conquests of intuition," and Weyl, "Logic is a hygiene we practice to keep our ideas healthy." As regards the context of discovery, the claim, much repeated in the recent literature, that thought processes in the mathematical sciences resemble those in the empirical sciences, would be entirely acceptable, did it not tend to suggest, what has not been established, that the thought processes of scientific discovery do not resemble those of artistic creation. For it seems, rather, that science is most clearly distinguishable from art in the context of justification, or rather, in *having* a context of justification, and that *there* rigorous proof is clearly distinguishable from systematic observation or controlled experiment.

Second and third, it must be acknowledged that as one moves from pure to more and more concretely applied mathematics, and from present-day to more and more remotely past mathematics, one eventually reaches a point where rigorous proof is less and less insistently demanded. The cases of applied and of past mathematics are related, inasmuch as past mathematics tended to be closer to applications in science. As regards rigor in proofs, the standards of Euler were far lower than those of Weierstrass. And it must be acknowledged that this is not because Euler was some mere obscurantist. Applications often cannot wait for rigorous proof. Indeed, it is a commonplace that the calculus, essential to modern science, would never have been developed if Newton and Leibniz had held themselves to a standard as high as that of Weierstrass.

What perhaps needs to be emphasized in the present climate of opinion is that, inversely, Weierstrass was not some mere pedant, and that applications often must wait for rigorous proof—or rather, new applications often emerge from the attempt to supply rigorous proof to the mathematics of old applications. Indeed, it is also a commonplace that the differential geometry and functional analysis applied in relativistic and quantum physics would never have been developed if Reimann and Hilbert had held themselves to a standard no higher than that of Euler. The period of about a century during which the standards of present-day pure mathematics have been in force is a small fraction of the history of civilization, but it is a largish fraction of the history of science. It is a period during which progress in mathematics has been cumulative, by addition of new theorems, while that of empirical science has been often radically revisionary, by amendment of old theories. This is an important difference, and one that will be overlooked by any approach that attends too exclusively to factors other than rigorous proof, to the most concrete applications, or to the most remote past.

Fourth, however, it must be acknowledged that, even in the context of publication in present-day pure mathematics, theorems and proofs are not everything. Principles resisting rigorous statement, conjectures resisting rigorous proof, with inductive, analogical, or heuristic arguments for them, and increasingly often computer verifications, simulations, or explorations, can also be found. In this regard the column of Wagon (1985) and its sequels must especially be cited. Enforcement of a high standard of rigor consists less of excluding such material from the literature than of assigning it a subordinate place, and especially of distinguishing it clearly from proofs of theorems. Wagon's column holds far more human interest than the typical theorem-proof-theorem-proof paper in a technical journal. So, too, do the studies, so common in the recent literature, emphasizing discovery, or applications, or the past. But scientifically, investigation of such factors belongs inevitably within the domain of the so-called human or soft studies, disciplines like psychology, sociology, history. To attend too exclusively to factors other than rigorous proof is to overlook the one factor that is susceptible to being investigated mathematically.

II

Classical logic is to be conceived of as having mathematical proof as its method of investigation. Its founders were all by education, and all but Russell by profession, mathematicians, and they never forgot it. This is so clear from their major works that it perhaps need not be argued here.

What do need clarification are the consequences of a conception of classical logic as a mathematical logic, a discipline similar in method to mathematical physics or mathematical economics (differing from them only in that its object is the very process of mathematical proof of which it and they are instances, giving it a self-referential character). One consequence is that, like mathematical physics, it deals with an idealization of reality. Its artificial languages are conspicuously simpler in grammar than natural languages, for example. No more than mathematical physics is mathematical logic to be condemned just for being unrealistic or idealized. Rather, like mathematical physics, it is to be evaluated by the success of its applications.

Another consequence is that, like mathematical economics, it provides descriptions rather than prescriptions, although, again like mathematical economics, its results can serve as minor premises in arguments with prescriptive major premises leading to prescriptive conclusions. Since the distinction between descriptive and prescriptive as applied to logic may be

unfamiliar, and since the claim is often repeated in the recent literature that logic is a normative, not a factual, discipline, some elaboration on this point is perhaps desirable.

Whenever a community has a practice, the project of developing a theory of it suggests itself. When the practice is one of evaluation, a distinction must be made between descriptive and prescriptive theories thereof. The former aims to describe explicitly what the community's implicit standards have been: the theory is itself evaluated by how well it agrees with the facts of the community's practice. The latter presumes to prescribe what the community's standards ought to be: the community's practice is evaluated by how well it agrees with the norms of the theory. Logic, according to almost any conception, is a theory dealing with standards of evaluation of deduction, much as linguistics deals with standards of evaluation of utterances. The distinction between descriptive and prescriptive is familiar in the case of linguistics: no one could confuse Chomsky with Fowler. It is not less important in the case of logic.

The familiar case of linguistics can help clarify a point about intuition important for logic. The data for descriptive theorizing consist of evaluations of members of the community whose evaluative practices are under investigation (e.g. "That's not good English"). These evaluations are of particular examples, not general rules. For even if it is supposed that members of the community, in making their evaluations, are implicitly following rules (e.g. even if it is supposed that grammatical rules are "psychologically real"), it cannot be supposed that they are capable of bringing these rules to explicit consciousness (e.g. how many lay native speakers of English can correctly state the phonetic rules for forming plurals?). Indeed, they are seldom acquainted with the technical jargon (e.g. "gerund") in which such rules are stated. Moreover, even the evaluations of particular examples are usually not stated in technical jargon: a negative evaluation will be primarily one of overall infelicity (e.g. "That isn't said"), and only secondarily if at all an indication of the dimension of infelicity involved (e.g. "ungrammaticality" as understood by professional linguists, as distinguished from overcomplexity, too obvious untruth or truth, etc.). While there is the least danger of bias when the data consist of spontaneous evaluations of spontaneous examples, in order to gather enough data elicited evaluations of contrived examples must be used. Indeed, theorists who are themselves members of the community (e.g. native English speakers investigating the grammar of English) often use their own impressions of the felicity or infelicity of particular examples as their main source of data. Such impressions are intuitions in an everyday sense, impressions of whose source and grounds one is unconscious. Intuitions in this sense are notably fallible and corrigible, especially in the case of a theorist out to establish a pet theory. If they conflict with the intuitions of other theorists, and especially if they conflict with the evaluations of unbiased lay members of the community, they must be retracted or at least restricted (e.g. to a dialect, rather than the whole language). To insist on them even given conflicting data (e.g. "People talk that way all the time, but it's wrong") would be to engage in prescriptive theorizing.

To apply this point to logic: descriptive logic is a branch of what has been called naturalized epistemology, epistemology so conceived that the epistemologist becomes a citizen of the scientific community, appealing to the results of the (other) sciences, and perhaps his or her own "intuitions" as a member of the community, in attempting to explain its practices from within; prescriptive logic is a branch of what may be called alienated epistemology, epistemology so conceived that the epistemologist remains a foreigner to the scientific community, evaluating its practices from without, attempting to provide a foundation for them or a critique of them, appealing perhaps to "intuitions" conceived of as fundamental critical insights from some extra-, supra-, or preter-scientific source of wisdom.

<center>III</center>

A conception of classical logic as a descriptive logic may be implicit in the conception of classical logic as a mathematical logic, as a branch of mathematics and hence of science; but no clear, detailed, explicit account of the distinction between descriptive and prescriptive logic, such as has just been attempted, is present in the major works of the founders. Nonetheless, in defense of the historical importance of such a conception, it can be argued that the work of the earlier founders is at least not explicitly prescriptive, and that that of the later founders is implicitly descriptive.

Such a claim may seem most doubtful in the case of Frege, whose work may seem so much a part of the process of increasing the standards of rigor in mathematics from Eulerian to Weierstrassian levels that was under way in his day. Yet even in his case half a dozen points can be cited in opposition to any interpretation according to which he was mainly concerned to prescribe to mathematicians a new notion of what constitutes full rigor.

First, the process of increasing the standards of rigor was itself less a matter of a new notion of what constitutes full rigor than of a new policy about how much rigor should be demanded. Mathematicians in Euler's day were less unaware of the fact that they were not demanding of themselves the highest standard of rigor, than unworried about it. Many quotations in the pertinent chapters of the standard history by Kline (1972), such as the following from d'Alembert (p. 619), illustrate this point: "More concern has been given to enlarging the building than to illuminating its entrance, to raising it higher than to giving proper strength to the foundations."

Second, in the process of increasing the standards of rigor, logicians, few in number and marginal in status, had little influence; and Frege seems to have been painfully aware of this fact. Hilbert is here the exception who proves the rule, since his great influence derived entirely from his great eminence in branches of mathematics other than logic.

Third, Frege himself, in the opening sections of the *Grundlagen* (1953), cites increasing the standards of rigor not as something that ought to be done but rather as something that is being done. He attempts to motivate his own project by appealing to an already accepted adage to the effect that in mathematics nothing capable of proof should be accepted without it.

Fourth, his own project was less concerned with securing the rigor of the steps that lead from one theorem to the next than with pushing back the starting point for theorem proving. Work of Dedekind and others was already achieving a reduction of the theory of real numbers to the theory of natural numbers, and an axiomatization of the latter theory; the work of Frege would achieve a reduction of the theory of natural numbers to the theory of classes, and an axiomatization of the theory of classes.

Fifth and sixth, he compares the relation of his ideography to natural language with the relation of an optical instrument to the naked eye. With this simile he simultaneously acknowledges the appropriateness of natural language for most mathematical work but insists on the importance of ideography for his own project. It is important there because, as the example of Euclid shows, it is hardest to uphold rigor when dealing with the most elementary material. All this, and the concluding sections of the *Grundlagen,* which advertise the *Grundgesetze,* suggest an interpretation of Frege's conception of classical logic less as a new standard of rigor than as a new instrument for checking that the old standard of rigor has really been upheld in material where appearances are least to be trusted: what occurs in the *Grundlagen* is a proof; what occurs in the *Grundgesetze* is a proof that it is a proof.

Russell can perhaps be interpreted in the same way, with the *Principles* standing to the *Grundlagen* as the *Principia* stands to the *Grundgesetze.* Peano cannot be so interpreted. He advocated the actual use of his pasigraphy by the mathematical community. But then he also advocated the use of his Latino (a rival of Esperanto) by the general public; so his attitude may be considered eccentric. After the time of Brouwer and the founding of the first nonclassical logic, the distinction between descriptive and prescriptive became clearer, especially in the work of the school of Hilbert.

Hilbert's Program involved conceding to the constructivists that much of classical mathematics is meaningless, but insisting that most of it is useful. To establish the latter point, it would be proved constructively that any statement that is meaningful constructively and provable classically is also provable constructively. To accomplish this it would suffice to prove constructively that no contradiction is provable classically. In accomplishing this, Russell's work is accepted as providing for classical mathematics (only) a "description of its methods—the good and the bad," in the words of the opening paragraph of a position paper by Hilbert's spokesman von Neumann (1983). Thus classical logic is not endorsed prescriptively as valid, but at most pragmatically as useful.

Hilbert's conception has been severely criticized by Kreisel (1983) on several grounds. One is that, by requiring that proofs about proofs, even about classical proofs, must be constructive proofs—a restriction perhaps appropriate for the one application Hilbert envisioned, but not for others—he created an obstacle to other applications. Classical logic is to be conceived of as useful for a variety of applications and not just a proof of noncontradiction, as with Hilbert, or a proof of reducibility (of number theory to class theory), as with Frege. The restriction Hilbert imposed on proof theory was indeed ignored in the development of model theory, in Skolem's Transfer Theorem, Gödel's Completeness Theorem, and Tarski's Compactness Theorem. It is perhaps only with Tarski (the present author's academic great-grandfather) and his conception of "the methodology of the deductive sciences" that one arrives at the conception of classical logic that has been expounded in the present paper.

Setting aside the question of extrinsic or historical importance, the intrinsic or scientific importance of such a conception of a "merely" descriptive logic might be questioned by prescriptive logicians. Though detailed consideration must be deferred, it may be mentioned that the tale of *Aladdin* warns us that it is unwise to trade old lamps for new until the comparative powers of the lamps have been thoroughly and exactly investigated. Thorough and exact investigation of the comparative powers of classical and constructivist logic and mathematics, in the work of Kreisel, Kleene, Feferman, Troelstra, Friedman, Simpson, and others, using classical logic as an idealized description of deductive practice in classical mathematics, has been and remains a main area of application of mathematical logic to epistemology.

IV

Russell's discovery of paradoxes and Gödel's discovery of incompleteness show that Frege's project and Hilbert's Program fail if considered solely and strictly in terms of their original intentions—though to have inspired the work of Russell and of Gödel is to have achieved more success than most intellectual enterprises can boast of. Hence classical logic must be evaluated by the success of other applications than those envisioned and intended by its original founders. In this connection it must be conceded that much of what is classified as mathematical logic today has no applications to epistemology. Some of

it (e.g. complexity theory, a branch of recursion theory) is applied to technology; most of it (e.g. degree theory, the main trunk of recursion theory) seems puristic. There remains the comparative investigation of the powers of the classical and the constructivist approaches, alluded to above. But perhaps the applications of most immediate pertinence to the work of mainstream mathematicians are those concerned with the scope and limits of classical or mainstream mathematics. The theorem of Gödel and Cohen,

(*) There is no proof or disproof in the classical logicians' sense, no formalized proof or disproof, of CH in ZFC.

or rather, the thesis inferred therefrom,

(#) There is no proof or disproof in the classical mathematicians' sense, no unformalized proof or disproof, of the continuum hypothesis in mainstream mathematics.

as well as similar work of Jensen, Solovay, and others, advises working mathematicians not to waste their time attempting to settle questions that are undecidable. One of Kreisel's complaints about Hilbert was that his "proof theory" could be more accurately entitled "provability theory," since it is mainly a theory about the existence, rather than other properties, of proofs. Theorems like (*) form so large a part of mathematical logic today that "unprovability theory" might be an even more accurate title.

Clearly the importance of such work in large part depends on the legitimacy of the inference to theses like (#). And this in turn clearly in large part depends on the appropriateness or adequacy of formalized ZFC as an idealization of unformalized Zermelo-Fraenkel set theory, and of the appropriateness or adequacy of Zermelo-Fraenkel set theory as an idealization of the starting point for deductions in mainstream mathematics. The former point seems unproblematic, since Zermelo-Fraenkel set theory is already semi-formal. The latter point may seem unproblematic, too, since Zermelo-Fraenkel set theory, or some variant, has a semi-official status in mainstream mathematics, appearing in the first volume of the encyclopedia of Bourbaki, and in popular textbooks like Halmos (1960).

Yet ever since the seminal paper of Benacerraf (1965), the philosophical literature has contained many expressions of puzzlement and perplexity over the role of set theory in mathematics. As Benacerraf observes, though Frege's reduction of numbers to classes or sets failed, others, due to Zermelo, von Neumann, and others, have succeeded. It seems that none of the various reductions of or surrogates for, say, the real numbers in set theory can be accepted as an account of what working mainstream mathematicians "really mean" by real numbers. Indeed, Halmos himself, in presenting material on one approach to reduction or system of surrogates, urges his readers to "learn it and forget it." And it does seem that the majority among the minority of working mainstream mathematicians who have learned it *have* forgotten it.

All this shows that a pure theory of sets like the Zermelo-Fraenkel theory is only an idealization of or approximation to mainstream or classical mathematics. It does not follow that there is any need for puzzlement or perplexity over the legitimacy of the inference from (*) to (#). On the contrary, what the reductions— whether Zermelo's or von Neumann's or another's— show is this: if one started not from the pure theory of sets but rather from a mixed theory of sets and real numbers, then in the mixed theory one could prove that the system of real numbers is isomorphic to a system of set-theoretic surrogates. The two systems share all properties that are preserved under isomorphism, and hence all properties of mathematical interest. Moreover, the properties that the real numbers could be proved to possess in the mixed theory are the same as the properties the surrogates could be proved to possess in the pure theory. Thus in investigations of what is provable or unprovable, the pure theory can be considered instead of the mixed theory: it is an adequate and appropriate approximation or idealization for purposes of "unprovability theory," even though not an accurate account of what anyone "really means."

Needless to say, the legitimacy of the inference from (*) to (#) also depends on the adequacy and appropriateness of classical logic as an idealization of or approximation to the working mainstream mathematical community's implicit standards for evaluating deductions. More precisely, the inference may be problematic unless what is deducible from what in the mathematicians' intuitive sense coincides with what is deducible from what in the logicians' technical sense. This thesis about deducibility might be called "Hilbert's Thesis." It parallels "Church's Thesis" about computability in the intuitive and in the technical sense. In defending this thesis, it must be conceded, indeed emphasized, that it only asserts that from the existence of a deduction or proof in the mathematicians' sense the existence of a deduction or proof in the logicians' sense may legitimately be inferred (and conversely, though the converse is not required for the legitimacy of the inference from (*) to (#)). As an account of other aspects or factors in the mathematicians' sense of proof, the logicians' sense of proof may be quite inaccurate, inadequate, and inappropriate. To illustrate this point it may be desirable to revert briefly to the topic, alluded to in an earlier section, of the influence of computers on mathematical practice. Indeed, it has often been suggested in the recent literature that this influence may be changing the mathematicians' sense of proof; and so it is perhaps necessary to emphasize that the status of Hilbert's Thesis is not being affected by any such changes as may be occurring.

One development has been the advent of "computer-assisted proofs," where calculations too long and complex for a human mathematician are done by machine, and what is published is the program for the machine and a report of its output, but not the whole output itself, which no human mathematician could check. While projects for having computers generate whole proofs have enjoyed only limited success in specialized contexts, allowing computers to assist with parts of proofs consisting of numerical or symbolic calculations has led to the announcement of solutions to some long-outstanding problems, notably the Four Color Problem in topology (for which see Appel and Haken (1986)). Most working mainstream mathematicians seem to accept that this problem has been solved, and speak of the "Four Color Theorem," not the "Four Color Conjecture." It can be—and has been, by Tymoczko and others—argued that the acceptance of "computer-assisted proofs" constitutes a change in the mathematicians' sense of what it is to give a proof. Indeed, it could be argued that rejection of "computer-assisted proofs" would also have constituted a change, since the mathematicians' sense of proof was formed before the advent of computers and before the question arose. What needs to be emphasized is that there has been no change affecting the status of Hilbert's Thesis. Indeed, the calculations left to the computer are much more like formalized proofs than are the conceptual parts of the proof undertaken by the human mathematician. Since computers are more reliable in checking long calculations than are human mathematicians in checking even short ones still done by hand, the quality of the evidence provided by the publication in the mathematical literature of a "proof" for the existence of a "proof" in the logicians' sense has not been decreased by the acceptance of assistance from machines.

Another development has been the advent, perhaps more as a theoretical possibility than a practical actuality, of "zero-knowledge proofs." Such a "proof" is a performance that makes it almost certain that the performer possesses a proof that, say, a certain natural number is composite rather than prime, but provides no information about such aspects of the proof as, say, what the factors of the natural number in question are. Renaissance algebraists used to give performances making it almost certain that they possessed algorithms for, say, solving quartic equations, without giving any information about most aspects of the algorithm. (They simply announced the solutions to many quartic equations posed to them as problems.) Such performances were not accepted as "giving an algorithm," and it may be doubted whether, if giving "zero-knowledge proofs" became common, it would be accepted as "giving a proof." But even if it were, and the mathematicians' sense of "proof were thereby changed, since the proofs the possession of which are made almost certain by such performances are formal proofs, or close to it, the status of Hilbert's Thesis would not be affected.

Even the best discussion by a commentator on the relation between proofs in the mathematicians' sense and proofs in the logicians' sense, such as that of Steiner (1975), will do less to convince a doubter of the truth of Hilbert's Thesis than will working through the formal proofs in *Principia Mathematica,* then the quasi-formal proofs (less and less formal in later and later volumes) in Bourbaki, and then informal proofs in the technical journals. For those not prepared to undertake this effort, the testimony of mathematicians who are not logicians, but who have examined the matter, may carry some weight.

Cohen was awarded the Fields Medal, the nearest equivalent to a Nobel Prize in mathematics, for his work; and surely not just for the technical virtuosity of the proof of (*), but also for the historical importance of (#). Thus the committee awarding this honor on behalf of the mathematical community, a committee composed entirely of nonlogicians, in a sense has evaluated the inference from (*) to (#) as legitimate. For that matter, Cohen himself, before and after his award-winning work, was an analyst rather than a logician, and presumably undertook to work on (*) because he was satisfied that from it (#) could legitimately be inferred. It is true that a few fanatical antilogicians among the category theorists have been overheard (by the present author among others) to say that the problem was not really solved until Cohen's work had been redone by one of them; but even this bizarre evaluation is an aspersion more on the proof of (*) than on the inference to (#). Most nonlogicians who have undertaken the effort to examine the matter, such as the prominent algebraic geometer Manin (1977), have been entirely positive in their evaluations.

The mathematical community as a whole has accepted the advice of the inference from (*) to (#) and abandoned any attempts to solve by mainstream methods a question mathematical logicians have inferred to be undecidable. The work of the few who are exceptions to this generalization bears all the signs of being the work of crackpots. Some mathematical logicians have continued to work on the undecidable questions, but not in the sense of trying to settle them by mainstream methods. Rather, their work is an exploration of proposed novel axioms. In this connection, the work of the self-styled "cabal" of Kechris, Martin, Moschovakis, Steel, Woodin, and associates requires special mention. The response of most working mathematicians to such work has in one way disappointed mathematical logicians. For the inference that a question is undecidable without novel axioms has very often been followed, not by an increased interest on the part of nonlogicians in the novel axioms, but rather by a decreased interest in the undecidable question, and even an evaluation of the whole branch of mathematics in which it arises as peripheral. Perhaps no different response is to be anticipated unless and until what has been done for the Continuum Hypothesis is done for something like the Reimann Conjecture. The disappointing current response, however, involves no aspersion whatsoever on Hilbert's Thesis. It is, if anything, an overenthusiastic acceptance of the advice of mathematical logicians that one will be wasting one's time if one attempts to solve certain questions. It is, if anything, a backhanded testimony to the success of classical logic in the kind of applications at which it aims according to the conception expounded in the present paper.

CONCLUSION (OR TRANSITION TO THE PROJECTED PART II)

Most commentators on nonclassical logics, from Haack (1974) onwards, divide them into the extraclassical, "supplements" to or "extensions" of classical logic, and the anticlassical, "rivals" or "alternatives" to classical logic. The latter classification, however, needs to be subdivided.

On the one hand, there are prescriptive critiques of logical practice, mainly addressed not to the world's dozens or scores of logicians, but rather to its thousands and myriads of mathematicians and scientists, or even its millions and billions of laypeople. On the other hand, there are descriptive critiques of logical theory. The traditional version of intuitionistic logic is an example of the first kind. The original version of relevance/relevant logic seems an example of the second kind.

According to the conception expounded in the present paper, not only the logics usually classified as extensions, but also those among the logics usually classified as alternatives that are of the first, prescriptive, practical kind, are directed toward aims different from those of classical logic. Only the alternatives of the second, descriptive, theoretical kind question the success of classical logic in achieving its aims as here conceived.

This is a distinction to which insufficient attention has perhaps been given in the literature. Detailed consideration of it in connection with specific nonclassical logics must be deferred. But clearly the conception expounded in the present paper leaves much room for nonclassical logics of the first kind, and little room for anticlassical logics of the second kind. When this point is accepted, the defense of classical logic is half (but as yet only half) complete.

REFERENCES

Appel, K. and Haken, W. (1986) "The Four Color Proof suffices," *Math-ematical Intelligencer* 8(1): 10–20.

Benacerraf, P. (1965) "What numbers could not be," reprinted in P. Benacerraf and H.Putnam (eds) *The Philosophy of Mathematics: Selected Readings,* 2nd edn, New York: Cambridge University Press, 1983, 272–94.

Bourbaki, N. (pseud.) (1964) *Elements de Mathématique, I: Théorie des Ensembles,* Paris: Hermann.

Frege, G. (1953) *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number,* trans, from the German by J.L. Austin, New York: Harper.

Haack, S. (1974) *Deviant Logic,* New York: Cambridge University Press.

Hadamard, J. (1954) *The Psychology of Invention in the Mathematical Field,* New York: Dover.

Halmos, P. (1960) *Naive Set Theory,* New York: Van Nostrand.

Kitcher, P. (1984) *The Nature of Mathematical Knowledge,* New York: Oxford University Press.

Kline, M. (1972) *Mathematical Thought from Ancient to Modern Times,* New York: Oxford University Press.

Kreisel, G. (1983) "Hilbert's Programme," revised version in P. Benacerraf and H.Putnam (eds) *The Philosophy of Mathematics: Selected Readings,* 2nd edn, New York: Cambridge University Press, 207–38.

Lakatos, I. (1976) *Proofs and Refutations,* New York: Cambridge University Press.

Manin, Y. (1977) *A Course in Mathematical Logic,* trans, from the Russian by N.Koblitz, New York: Springer.

von Neumann, J. (1983) "Formalist foundations of mathematics," trans, from the German by E.Putnam and G.Massey, in P.Benacerraf and H.Putnam (eds) *The Philosophy of Mathematics: Selected Readings,* 2nd edn, New York: Cambridge University Press, 61–5.

Polya, G. (1954) *Induction and Analogy in Mathematics, 2* vols, Princeton, NJ: Princeton University Press.

Steiner, M. (1975) *Mathematical Knowledge,* Ithaca, NY: Cornell University Press.

Tymoczko, T. (ed.) (1987) *New Directions in the Philosophy of Math-ematics,* Boston, MA: Birkhauser.

Wagon, S. (1985) "The evidence," *Mathematical Intelligencer* 7 (1): 72–6.

2
PROOFS AND EPISTEMIC STRUCTURE
Glen Helman

SUMMARY

The point of application of logic to epistemology is usually thought to lie in the concept of an argument. Here an attempt is made to deepen the connection between the two fields by means of the concept of a proof. Specifically, the structure of proofs is compared with the structure of the information employed in assessing the justification of belief.

Arguments serve epistemic assessment by presenting information about the believer's reasoning, but the information required for judging the belief cannot be exhausted by argument. The epistemic role of an argument should instead be understood by analogy with pointing: an argument serves to *indicate* a justification for its conclusion. The importance of information both about global features of the current body of belief and about the sources of the belief whose justification is being assessed suggest that what an argument indicates is a fragment of an epistemic biography. The biographical fragments associated with various contemporary beliefs are related, but these relations are neither genetic nor constitutive; in particular, we should not suppose that some are constructed out of others.

Biographical fragments are objects associated with arguments on epistemic grounds. On the side of logic, a relation of equivalence holding between proofs of the same conclusion by essentially the same means will induce equivalence classes which may be thought of as abstract proofs. These abstract proofs form a domain structured by operations corresponding to rules of proof. The structure of this domain may be interpreted in part by operations familiar from theories of constructions (pairing, functional abstraction, etc.). But we can expect no constructive interpretation of the structure corresponding to a full system of rules for classical logic, and we are left to look elsewhere for an account of the domain of abstract proofs.

So both epistemology and logic can profit from an identification of abstract proofs with the biographical fragments needed for the assessment of belief. Epistemology gains a way of articulating (some of) the relations among these fragments while logic gains an interpretation of the equivalence of proofs and rules of proof. But a variety of issues must be faced before the identification can be secured. Few of these issues are settled here, but they are explored in a number of directions. One of these concerns the epistemic role of the logical constants and the significance that may be given to the validity of arguments.

The application of logic to epistemology is usually understood to have its locus in the concept of an argument. Arguments are used in the justification of belief and they are open to logical evaluation. The applicability of logic is secured if we have an account of this evaluation which shows its importance for epistemic assessment. Semantic accounts of deductive validity do just this, and the various approaches to inductive argumentation attempt something similar.

Still, we may hope for a deeper connection between the two fields. Each enriches and extends the concept of an argument for its own purposes. The needs of epistemic assessment are not met by the simple citation of certain beliefs as premises. We also need information about global features of the whole body of belief and about the specific way a belief comes to be held. So, although an argument may be cited in justifying a belief, the justification cannot consist of the argument alone. For its part, logic extends the concept of an argument to that of a proof. Proofs provide information about the ways that conclusions are reached from premises that is unnecessary if our only interest is in the validity of arguments. However, the study of proof has been fruitful enough to give its objects an independent standing, and it has led to suggestions for relations of reduction and equivalence which structure the domain of proofs.

One way to deepen the connection between logic and epistemology is to compare the structure of proofs to the structure of the information employed in assessing the justification of belief. In outline, such a comparison may proceed as follows. The concepts of a simple argument and of a proof are each subsumed by the more general concept of a possibly compound argument, an argument which may have the internal structure of a proof while containing nondeductive steps. The epistemic role of such an argument may be understood by analogy with the use of descriptions in ostension: it helps to identify a justification for its conclusion by reference to justifications of its ultimate premises. The possibility of such relative identification implies that justification has a structure. It is this structure that we may compare with the algebraic structure which is exhibited by a system of proofs modulo a relation of equivalence between proofs. If the structure of justification is

the same as that found in proofs, a description of justificatory structure can serve as an interpretation of rules of proof and of the equivalence of proofs formed using these rules. It can thus provide an epistemic foundation for logic.

In attempting to fill out this sketch, we will consider first the epistemic function of argument and then the structure exposed in proofs by a relation of proof equivalence. Finally, we will survey the issues that must be faced if we are to ascribe the structure of proofs to features of epistemic justification and employ this structure in an interpretation of proofs.

I.
ARGUMENT AND THE JUSTIFICATION OF BELIEF

The function of arguments

When an argument is used to justify a belief, it serves to present information which is needed to assess the beliefs justification. And it is to the nature of this information that we must look to find the epistemic function of arguments. In the discussion to follow, we will presuppose a concept of justified belief as distinct from knowledge. No attempt need be made to settle the relation between the two; we need only assume the difference between these concepts to lie in the range of information needed for their application. A belief may be recognized as justified on the basis of a proper part of the information that would be needed to recognize it as knowledge. The information needed is roughly that part of the information needed for the assessment of knowledge for which our primary source is the believer's testimony (though there is no need to assume that this testimony is incorrigible or that it concerns internal states and processes). We will also assume that the partial information needed to assess the justification of a belief determines what further information is relevant to assessing it as knowledge. This assumption will enable us to focus on the structure of justification, with the expectation that this is mirrored by the structure of grounds of knowledge.

The information needed to assess the justification of a belief includes both the evidence for the belief and the use of this evidence in reasoning.[1] In the absence of either sufficient evidence or effective use of this evidence, a belief may simply be a guess. Accordingly, we may drop the distinction between evidence and its employment and speak simply of reasoning, understanding this to include any starting point in evidence. It is information about reasoning (in this sense) that is the basis for the assessment of justification and is the believer's contribution to the assessment of knowledge.

Information about one's reasoning is often presented in an argument. Of course, a single-step argument from one or more premises may not suffice to make clear one's reasoning for a belief. The obscurity that remains can lie in either the reasoning for a premise or in the step from premises to conclusion. In both cases, the simple argument may be elaborated to a compound argument either by presenting an argument for a premise of the original argument or by articulating the step from premises to conclusion by inserting one or more intermediate conclusions. The full range of arguments (in the sense in which we will use the term here) includes such compounds and also compounds which involve the use of hypothetical assumptions or generalization on singular terms. As a result, proofs in a system of natural deduction will count as arguments formally presented though, of course, arguments will not generally proceed according to formal rules and will usually contain steps that are not deductive.

Now, how do arguments serve to inform others of one's reasoning? The simplest account would be that one's reasoning for a given belief can be fully represented, in principle at least, by a certain completely elaborated argument. The same reasoning could be explained in practice by one of the many less elaborate approximations to this ideal argument, and such explanations could be understood simply as cases of ellipsis. However, this account of the function arguments entails a very strong sort of epistemic atomism. If an argument is completely elaborated, its premises and individual steps must be elementary in the sense that no further elaboration is ever needed. In order for every argument to have a complete elaboration, there must be a supply of elementary premises and elementary steps of argument which suffice to represent in full the reasoning for any belief, justified or not. There are familiar and serious difficulties with such a view. Some of these will be cited later, but we will assume now that they are clear enough to render implausible any account of argument as an elliptical presentation of reasoning.

Although we may not, even in theory, take reasoning to be completely represented by argument, we must still account for the fact that one's reasoning can be presented in practice by arguments of varying degrees of elaboration. There is an analogy here with the use of descriptions in ostension. While it is generally impossible to replace a demonstrative uniformly by a description with no demonstrative components, descriptions can play an important role in helping us indicate the object referred to. Moreover, when a simple description fails, we are led to replace it by a more elaborate one, the simple "that book" becoming, for example, "that red book on the far right of the top shelf."

Let us follow this analogy and speak of an argument as serving to *indicate* reasoning. A simple argument indicates the reasoning for its conclusion by reference to the reasoning for its premises. It can succeed in this only if the reasoning for its premises is clear without aid of argument. What we have for each premise is only the analog of a pure demonstrative inflected

by nothing but a statement of the premise. And the way that the reasoning for the conclusion is related to that for the premises must also be clear. Obscurity of either sort may be attacked by elaborating the argument, thereby providing a more detailed description of the reasoning. Demonstrative elements will remain in the premises and the individual steps of the argument; but, once an argument is elaborated in such a way that the referents of all its demonstrative elements are clear, the reasoning for its conclusion will have been identified.

Epistemic biography

But what is it that we are able to identify in this way? The most that can be said in general is that one's reasoning is one's biography from an epistemic point of view. The need to cast our net this wide can be seen in three steps. First, we must grant that, in assessing a beliefs justification, information about the other currently held beliefs is needed both to distinguish between evidential and nonevidential beliefs and to assess the strength of nondeductive arguments. The next step to epistemic biography is the observation that even the whole body of current belief cannot be assessed in isolation but must be judged with respect to its past. The final step is to note that the information needed about past belief cannot be limited to the features of one preceding body of belief; that is, even changes from one body to another can be assessed only in the light of beliefs at other times. The first of these steps can be based on standard arguments for the contextual character of both the evidential status of beliefs and the strength of nondeductive inference.[2] The second and third steps are less securely part of common philosophical wisdom and deserve further comment.

Both concern the need for information about the past. Without information about previous history, the only judgments we can make about someone's beliefs concern such features as consistency, simplicity, and the tightness of explanatory connections. Let us group these features together under the label of "coherence." Although it may be less generally granted, it also seems true that one may not adopt just any coherent set of beliefs, even one that is maximally coherent. Coherence is a much richer notion than consistency, but it still leaves much room for variety in belief. And it may not be reasonable for me to now hold one quite coherent set of beliefs if I have just held another very different one. Of course, radical changes in belief may be reasonable; but arbitrary changes are not, even when the resulting body of belief is entirely coherent. And, given this, the assessment of my current beliefs depends in part on information about those in my past.

The final step to epistemic biography is the observation that this information about the past cannot be limited to information about a single preceding set of beliefs. Such information would be sufficient to specify fully a change in belief, but even a change cannot be assessed on internal grounds alone. We need information about both preceding and intervening stages, about both the historical background for the change and about its course. Different changes will be justified depending on whether a given body of belief is or is not justified with respect to its past. A goal of conservation that we may wish to govern any change from a justified set of beliefs may not be appropriate when the initial beliefs are unjustified; stubbornness in error does not have the value of resoluteness in justice. But we can assess the justification of the body of belief preceding the change only if we have information about its history. Information about the course of a change is also important. A change that seems reasonable when assessed only in terms of its beginning and end will look very different if closer inspection shows it to be the net result of a number of entirely arbitrary steps.

Now, the biographical information needed to assess belief is not only extensive in the ways just surveyed; it is also rich. A mere chronology of beliefs held, however complete it is, will not be enough. We need to have an account not only of which changes have occurred but also of why they have occurred. In assessing a change, we cannot confine ourselves to comparing the contents of the resulting body of belief with those in the past. We must go on to ask, for example, whether a new belief is added as a perceptual judgment, an inference, or a guess. And there is no reason to think that information about which beliefs are held before and after the change will settle these questions. Since people have beliefs about their reasoning, some of the beliefs produced by the change may concern the reasons for its occurrence. It is also true that the believer's testimony will always be our primary source of information. But, in the end, it is what the reason for the change is, not what it is believed to be, that is crucial for epistemic assessment.

While the assessment of an individual belief requires, in principle, information about the whole of the believer's epistemic biography, this information need not be exhaustive. When we consider the contemporary context of a belief, our questions concern only the truth of certain generalizations about this context; we may need to know, for example, that there is no belief that defeats a certain presumption. Such information need not include full details about which beliefs are held. The same will be true of information needed about past beliefs and about the processes and procedures by which the beliefs have developed. A figure may help here. To assess a belief, we must understand its place in the wider society of belief, and for this it is not enough to know a family tree of premises on which it is based. We need information about the whole body of belief and its history, but we do not need to know these things in full detail.

An argument, then, serves to indicate certain biographical information concerning the belief that is its conclusion, and it does so by reference to similar information concerning its premises. Such relative identification presupposes relations among these collections of information. That there should be such relations is not surprising since the information relevant to the

premises and the conclusion is all drawn from the same biography and, moreover, from a biography that is structured by an account of the development of belief. As an example of the sort of relation of collections of information we can expect to find, consider a prediction of a future event which arises by generalization on past experience. An account of the prediction's origin will be part of the believer's overall epistemic biography and will serve to tie the biographical information relevant to the prediction to the information relevant to beliefs about the past. Furthermore, this kind of inductive connection is sufficiently common for the believer to hope to indicate it simply by offering an argument for the belief about the future with particular beliefs about the past as its premises.

In addition to connections among fragments of biographical information which reflect the genesis of various beliefs, there are connections which derive from the fact that epistemic biography, like any biography, is a theoretical construction. Some features of the structure of epistemic biography will be artifacts of general constraints on theories of this sort. For example, if we ascribe to someone a belief in a conjunction, we will generally also ascribe beliefs in the conjuncts; and the portions of the constructed biography relevant to each of the three beliefs will stand in a particular relation as a result. We need not suppose that constitutive connections of this sort are sharply distinguished from the genetic connections already noted, and the details of the interrelation of these two sorts of connection can be left unsettled here. The point is simply that we can expect epistemic biography to be subject to the general principle that both the properties of the subject matter and the exigencies of theoretical construction affect the structure of theoretical understanding and the structure of the knowledge it produces.

A final caveat is in order. While genetic and constitutive factors may lie behind the relations among partial biographies, these relations should not themselves be thought of as either genetic or constitutive. In the first place, beliefs in the premises and conclusion of a given argument are contemporary and are each assessed with respect to the same body of belief in light of the same history. So the bodies of information relevant to the assessment of the individual beliefs are not ordered temporally although they may stand in atemporal relations which reflect the genetic information they contain. Second, while we ascribe some beliefs because we ascribe others and thus constitute epistemic biography in part by stipulation, we do not thereby construct some partial biographies out of others. All partial biographies stand on the same level of abstraction as aspects of the full epistemic biography seen from different perspectives. These aspects may be related in ways which reflect the constitution of the biography as a whole, but such relations are themselves no more constitutive than are the relations among different views of a physical structure.

There are a number of further issues regarding the structure of epistemic biography that are relevant to the function of argument and the application of logic, but these will become apparent only when we have considered the structure of proofs. So let us recall that the sort of argument we have been considering can be presented formally as a proof in a natural deduction system, possibly with nondeductive steps, and go on for now to look at proofs from a formal point of view.

II.
THE STRUCTURE OF PROOFS

Although there seems to be an intuitive concept of equivalence applying to proofs—one which is employed, for example, in judging the originality of mathematical work—most technical developments of the idea have been by-products of other concerns. The two principal sources are the theory of normalization for natural deduction proofs and various attempts to find in proofs yet another instance of the general constructions of category theory.[3] Our purposes here will be served by considering only a few general features of proof equivalence, which may be motivated directly without developing either a theory of normalization or the relevant ideas of category theory. The substance of the discussion will not be new, and it will stop far short of a full development of the concept of proof equivalence.[4] We will sketch the abstract structure of proofs only in the detail needed to permit comparison with the structure of epistemic biography.

We will assume that the proofs take the form of derivations in a system of natural deduction with the usual introduction and elimination rules for logical constants, but with one stipulation about the form of these derivations which will be a help in describing their structure: a proof may contain distinct undischarged assumptions of a given formula, and such assumptions can be reidentified in different proofs. This assumption has the consequence that two proofs may differ solely in the way their undischarged assumptions match those of some third proof. In particular, for each formula A, there may be many (and we will assume that there are infinitely many) degenerate proofs each of which consists solely of an undischarged assumption of A. And these proofs will be spoken of as the *assumptions* of A. Any undischarged assumption of A in a more complex proof will be matched with one of these and regarded as an occurrence of it as a subproof. Discharged assumptions in different proofs will not be identified, so distinctions among proofs may disappear with the discharge of assumptions.[5] Under these conventions, undischarged assumptions are entirely analogous to free variables; and discharged assumptions have a status analogous to that of bound variables when we do not distinguish formulas which differ only in their choice of bound variables. In the following, "D" (perhaps with a subscript or primes) will serve as a variable ranging over proofs, with

$$\frac{D}{A}$$

(and, in linear form, "D:*A*") used to indicate the conclusion. Variables for assumptions will take the form *"[A]"*, with the primes orsubscripts added to the brackets when we wish to consider distinctassumptions of the same formula.

Abstract proofs

We may assume that proof equivalence is restricted to proofs of the same conclusion. Given this restriction, each equivalence class of proofs modulo proof equivalence will have a definite conclusion. Since proofs may contain undischarged assumptions (which may be thought of as placeholders for accepted premises or nonlogical axioms), there will be proofs (indeed, infinitely many) for each formula. Our notation for the relation of proof equivalence will be "~," and both "D/~" and boldface "**D**" will be used for the equivalence class of D (so that , or **D=E**, if and only if). An equivalence class of proofs with conclusion *A* can be thought of as an *abstract* proof of *A,* a way of proving *A* which has been abstracted from its various syntactic presentations.

Proofs can be combined by using a proof D_1 of a formula *A* to support an assumption of *A* in another proof D_2. In doing this, we elaborate an argument by offering an argument for one of its premises. This operation can be thought of syntactically as the substitution of a proof for an assumption of another proof; in the following,

$$\begin{array}{c} D_1 \\ [A] \\ D_2 \end{array}$$

(linearly, "$D_2(D_1/[A])$") will serve as notation for D_2 with D_1:*A* substituted for *[A]*. Let us assume that substitution preserves equivalence both of proofs which are substituted and of proofs in which substitution occurs. That is, we will adopt the following principles.

1 *Functionality:* If D_1~D_2, then $D(D_1/[A])$~$D(D_2/[A])$, i.e.

$$\begin{array}{ccc} D_1 & & D_2 \\ [A] & & [A] \\ D & \sim & D. \end{array}$$

2 *Generality:* If D_1~D_2, then $D_1(D/[A])$~$D_2(D/[A])$, i.e.

$$\begin{array}{ccc} D & & D \\ [A] & & [A] \\ D_1 & \sim & D_2. \end{array}$$

The first tells us that if we substitute each of two equivalent proofs of a formula *A* for an assumption of *A* in some proof, then the results are equivalent. According to the second, if we substitute a proof of a formula *A* for an assumption of *A* in each of two equivalent proofs, then the results are equivalent. The two together enable us to define an operation

$$\mathbf{D(D'/[A]) = D(D'/[A])/\sim}$$

of substitution for abstract proofs. To apply this function we choose concrete representatives D and D of the abstract proofs **D** and **D** , carry out ordinary concrete substitution and then abstract again by ~.

The consequences of these assumptions for the abstract structure of proofs are best seen in a somewhat different form. A concrete context D(−/[A]) in which substitution can take place determines a function which maps each proof D :*A* to the result D(D /[A]) of substituting it in this context. It also determines an abstract context **D**(−/[A]) and a corresponding function defined on abstract proofs of *A*. Indeed, the assumptions above are equivalent to the claim that, for any abstract proof **D** and assumption *[A],* there is a function h(*[A]*, **D**) on abstract proofs of *A* such that

$$h_{[A], \mathbf{D}}(\mathbf{D'}) = D(D'/[A])/\sim.$$

The function h(*[A],* **D**) can be seen as representing the way that the substituted abstract proof is used in the abstract context **D** (−/[A]).

The operation of substitution and the principles of functionality and generality can be extended to the case of several assumptions, justifying the definition of many-place functions h($[A_1]_1$,…, $[An]n,$ **D**). The upshot is that for each argument with premises A_1,…, A_n and conclusion *B,* there is a collection of *n*-place functions each of which applies to any sequence of

abstract proofs of $A_1,\ldots, A_n$ to yield an abstract proof of *B*. Each such function represents, in abstraction, the effect of a proof when it is combined with proofs of certain of its assumptions. We will refer to these functions also as abstract proofs and distinguish abstract proofs for arguments from abstract proofs of sentences by referring to the two sorts as *hypothetical* and *categorical* abstract proofs, respectively.

This gives us the basic framework of the abstract structure of proofs, a family of domains together with a family of function spaces defined over them. Each of the domains contains the categorical abstract proofs of a given formula, and each of the function spaces contains the hypothetical abstract proofs for a given argument. Additional structure will reflect the various ways proofs are compounded out of subproofs.

Rules of proof

Let us first see how rules appear in the abstract structure of proofs. The key assumption here is that a rule preserves equivalence for proofs to which it is applied. This assumption has the consequence that an operation on abstract proofs is associated with each rule; however, the form that this operation takes will depend on whether or not the rule discharges assumptions. A rule which does not discharge assumptions (e.g. the introduction rule for conjunction) is a scheme which draws a conclusion from premises of a certain related form. We can find an associated operation on abstract proofs for a particular choice of premises and conclusion by considering a proof in which the rule is applied directly to assumptions. For example, given proofs D_1:*A* and D_2:*B,* let &I(D_1, D_2) be the proof of *A&B* by &-introduction on the conclusions of D_1 and D_2. Any proof &I(D_1, D_2) can be thought of as the result of substituting D_1:A and D_2:B in a proof &I*([A], [B])* that consists of an application of &I to assumptions of *A* and *B*.

The principle of functionality thus enables us to define an operation **&I** on equivalence classes by

$$\mathbf{\&I(D_1, D_2)} = \&I(D_1, D_2)/ \sim.$$

This operation will apply to any pair of categorical abstract proofs, and its restriction to proofs of particular formulas *A* and *B* will be a hypothetical abstract proof for the argument from *A* and *B* to *A&B*.

A rule which discharges assumptions in proofs of one or more of its premises also has an associated operation. However, this operation does not apply to categorical abstract proofs of the premises in question but instead to hypothetical abstract proofs which represent the ways in which the discharged assumptions are used in proofs of the premises. For example, an instance of the introduction rule for the conditional applies to a proof D:*B,* discharging an assumption *[A]* to yield a proof →I(*[A]*, D) of *A* →*B.* The associated operation applies to an abstract hypothetical proof for the argument from *A* to *B* to yield an abstract categorical proof of *A* →*B.* In the case of rules which do not discharge assumptions, the existence of corresponding operations was a consequence of the principle of functionality; but, for the case of rules which do discharge assumptions, we must make specific assumptions about the conditions under which the proofs that result from their application are equivalent.

In the case of →I, it is natural to assume that →I(*[A]₁*, D_1) and →I(*[A]₂*, D_2) are equivalent if D_1 and D_2 use the assumptions *[A]₁* and *[A]₂*, respectively, in equivalent ways. That amounts to the assumption that →I(*[A]₁*, D_1) and →I(*[A]₂*, D_2) are equivalent if h(*[A]₁*, $\mathbf{D_1}$)=h(*[A]₂*, $\mathbf{D_2}$). Given the principle of generality, we can get the same effect by the following principle of proof equivalence.[6]

Replacement in → I: If $D_1 \sim D_2$,

then $\rightarrow I([A], D_1) \sim \rightarrow I([A], D_2)$.

An operation **→I** on hypothetical abstract proofs may then be defined by[7]

$$\rightarrow \mathbf{I(h_{[A],D})} = \rightarrow I([A],D)/ \sim.$$

Operations on abstract proofs can be associated with other assumption-discharging rules in a similar way.

The rules for quantifiers which generalize on proofs, like ∀ and ∃E introduce a still different sort of structure. For these we need to consider entities that might be called *general* abstract proofs, both categorical and hypothetical. In the simplest case, an abstract proof of *A* general with respect to a variable *x* assigns to each term t an abstract proof of *A (t/x),* the result of substituting t for *x* in *A*. This sort of abstract proof can be thought of as representing a way that a proof employs the denotation of a term, and a rule (like ∀ or ∃E which enables us to generalize on parametric terms will be associated with an operation that applies to general abstract proofs. Since this enrichment of the structure contributes nothing especially significant for our purposes here, we can continue to focus solely on the case of prepositional logic.

The principles we have considered so far concern only connections between the equivalence of component proofs and the equivalence of proofs as a whole. They do not imply any categorical claims about the conditions under which proofs are equivalent; indeed, they hold for the relation of syntactic identity in addition to many broader relations. More substantial principles are clearly of interest for a full account of the equivalence of proofs, but their interest for the project at hand lies

primarily in the form we can expect them to take; and this can be seen in a few simple examples. One natural equivalence concerning the rules for conjunction is the following: $\&E_i(\&I(D_1, D_2)) \sim D_i$ for $i=1, 2$, i.e.

$$\&E_i \cfrac{\&I \cfrac{\begin{array}{cc} D_1 & D_2 \\ A_1 & A_2 \end{array}}{A_1 \& A_2}}{A_i} \sim \begin{array}{c} D_i \\ A_i \end{array}$$

(where $\&Ei$ is the case of &-elimination that derives A_i from $A_1 \& A_2$). That is, if we derive a conjunction $A\&B$ from proofs of A and B individually by the introduction rule and then go on to conclude one of the two by the elimination rule, the result is equivalent to our original proof of the individual conjunct. The elimination rule undoes the effect of the introduction rule up to proof equivalence. Read right to left, this is what Prawitz (1971) calls a reduction law for conjunction in his account of normalization for natural deduction proofs. Another natural principle tells us that introduction undoes the effect of elimination: $\&I(\&E_1 (D), \&E_2(D)) \sim D$ (where $D:A\&B$ for some A and B), i.e.

$$\&I \cfrac{\&E_1 \cfrac{\begin{array}{c} D \\ A\&B \end{array}}{A} \quad \&E_2 \cfrac{\begin{array}{c} D \\ A\&B \end{array}}{B}}{A\&B} \sim \begin{array}{c} D \\ A\&B. \end{array}$$

Read right to left, this is what Prawitz calls a law of expansion. It tells us that any proof of a conjunction is equivalent to a proof by &I (where &I is applied to the results of applying the two forms of &E to the original proof).

When we abstract, these and similar principles become equations between abstract proofs which state algebraic laws for the operations corresponding to rules of proof. To see the form these laws take in the case of conjunction, let us define a function **&E** from abstract proofs of conjunctions to pairs of abstract proofs by

$$\mathbf{\&E(D) = (\&E_1(D), \&E_2(D)).}$$

In the example above, the first principle enables us to recover (up to equivalence) the proofs to which &I has been applied by applying $\&E_1$ and $\&E_2$. It follows that **&I** is a one-to-one function on pairs of abstract proofs and, moreover, that the operation &E is a left inverse for **&I** (i.e. we have **&E(&I(D$_1$, D$_2$))=(D$_1$, D$_2$)).** The second principle can be read to say that the operation **&I** is a function *onto* the abstract proofs whose conclusions are conjunctions (since every such abstract proof contains a concrete representative proof by &I) and, indeed, that **&E** is a right inverse for **&I**. The two principles together then tell us that the operations corresponding to conjunction rules are mutual inverses and provide a one-to-one correspondence between pairs of abstract proofs and abstract proofs of conjunctions. Abstract proofs of conjunctions can thus be thought of as coding pairs of proofs in the way numbers can code pairs of numbers.

Something similar happens with   . We have assumed that   I is an operation that carries hypothetical abstract proofs to abstract proofs of conditionals. Natural principles of proof equivalence can be stated which will characterize it as 1–1 and onto. Abstract proofs of conditionals can then be understood to code hypothetical abstract proofs, a proof of A B coding a function from proofs of A to proofs of B. Although natural, these principles can be a matter of controversy, as can those for the &-rules. And the situation is even less settled in the case of other logical constants.[8] But the key point for our purposes is that the principles which might be accepted for specific rules of proof can be given the form of algebraic laws for the corresponding operations.

Let us pause again now to collect what has been said about the structure of abstract proofs. It can be thought of as a many-sorted algebra with a domain assigned to each formula and each argument and with various operations, each defined for certain domains. The laws which give the specific character of the structure are of three sorts. First, there are principles which tell us that the domains associated with arguments consist of functions. That is, among the operations of the structure are application operations (with varying numbers of places) which apply the hypothetical abstract proofs associated with arguments to abstract proofs of the premises of these arguments. Along with these operations come principles of identity for hypothetical abstract proofs. These are consequences of an extensional criterion of identity which states that two proofs for an argument are identical if they yield the same proof of its conclusion no matter what proofs of its premises they are applied to.

A second sort of law for the structure implies the existence of certain hypothetical abstract proofs. For example, abstract proofs for arguments from A to B and from B to C may be composed as an abstract proof for the argument from A to C. These laws follow from the fact that there is a hypothetical abstract proof h($[A_1]_1$, …, $[A_n]_n$, D) associated with each concrete proof D and list of assumptions $[A_1]1,…, [A_n]_n$. Although these laws can be given an algebraic form, they can be stated most simply as asserting closure of the structure under a form of definition for which concrete natural deduction proofs provide the notation.

Finally, as a third group of laws, we have the algebraic principles which state relations among the distinguished operations corresponding to rules of proof (such as the fact that **&I** and **&E** are mutual inverses). These provide filling for the framework outlined in laws of the first two sorts and reflect the specific features of a given system of proof.

It is important to remember that we may take hypothetical abstract proofs to be functions individuated extensionally only if we are granted certain assumptions about the existence of categorical abstract proofs. In particular, a substantial supply of such proofs is needed, and this supply must include proofs that will count as nondeductive. This can be seen most easily with regard to negation. A natural choice of principles of equivalence for negation rules yields a one-to-one correspondence on the abstract level between the proofs of negations ¬A and the proofs of arguments from *A* to a contradictory conclusion   , the reductions of *A* to absurdity. Given this correspondence, there can be distinct abstract proofs of a given negation ¬*A* only if we have a variety of abstract proofs for reductions of *A* to absurdity, and we can have this variety only if we admit a number of abstract proofs of the logical falsehood   . This requirement is satisfied by the structure of abstract proofs as it has been described here because among the proofs from which we abstract are those containing undischarged assumptions, giving us an abstract proof corresponding to each assumption [   ] of absurdity.[9]

The abstract proofs that do count as deductive will include at least one categorical proof for each logical truth and at least one hypothetical proof for each valid argument, but not all abstract proofs for logical truths and valid arguments will count as deductive. The purely deductive abstract proofs are the categorical abstract proofs corresponding to concrete proofs all of whose assumptions are discharged and any hypothetical abstract proof h($[A_1]_1$, …, $[A_n]_n$, **D**) corresponding to a concrete proof D whose undischarged assumptions are among the assumptions $[A_1]_1$, …, $[A_n]_n$. Each of these abstract proofs can be identified uniquely by reference to the basic operations of the structure which correspond to rules of proof, and therefore each has a distinguished status comparable with the status of the basic operations. On the other hand, the various assumptions of a given sentence are not distinguished by the structure; a mapping can preserve the structure while carrying one assumption into another.[10]

Interpreting the abstract structure

Once we have such an algebraic description of the structure of proofs under proof equivalence, we can cut its tie to a specific realization by abstract proofs. It is then possible to work in the other direction, interpreting rules of proof by the operations of the structure and attempting to assign each proof an object in the structure as its denotation. It will be possible to produce this assignment if we have stated enough closure conditions on the structure. And, if we have stated a large enough body of algebraic laws, any two equivalent proofs will receive the same denotation for any assignment to their undischarged assumptions. Further, since proofs under the relation of proof equivalence provide one instance of the structure abstracted from them, any proofs that always receive the same denotation must be equivalent. So, if we have described the algebraic structure we find in abstract proofs in sufficient detail, this structure, once it is abstracted, can be used to characterize the relation of proof equivalence.

Of course, for such an account of proof equivalence to be enlightening, we need not only an independent description of the abstract structure but also an independent understanding of its significance. In the case of intuitionistic logic, this can be provided by a theory of constructions. Many such theories have been proposed in the last fifty years, but all derive from Brouwer's idea that mathematical language is an instrument for communicating mathematical constructions. In Heyting's formulation of this, a mathematical assertion reports either the completion of a mathematical construction or the possession of a general method for such construction. The construction or method itself is the proof of the assertion in an abstract sense; a verbal proof provides a description of it. These ideas can be developed as an account of the significance of logical connectives by associating with each connective a particular relation holding between the constructions for compound sentences formed using the connective and the constructions for the components of these sentences. According to one common way of doing this (which is close to Heyting's informal accounts of the connectives) the construction associated with a conjunction is an ordered pair of constructions and the construction associated with a conditional is a function from constructions to constructions.[11]

So understood, constructions are mathematical objects which comprise a structure of the general sort needed to interpret proof equivalence in the way outlined above. However, this way of understanding the abstract structure of proofs is not open for classical logic. The extra content of the notion of a construction over that of an abstract proof is to be found in a close tie between the constructions associated with sentences and those associated with terms. Together they form a single system of mathematical objects. But when the structure of proofs contains an operation corresponding to classical indirect proof, this tie is loosened. An indirect proof is indefinite but not because it specifies a ground for its conclusion in an indefinite way, for there need not exist any alternative definite construction of this conclusion. The indefiniteness lies rather in the ground itself, which is without definite ties to the constructions which may be given for terms. And any concept of "construction" sufficiently liberal to encompass such indefinite grounds will be empty of the special content suggested by Brouwer's ideas.

A second problem with the interpretation of abstract proofs as constructions is that it arranges the domains of proofs for various formulas in a hierarchy of increasing abstraction—as spaces of pairs, function spaces, and so on—according to the logical forms of their conclusions. This makes it difficult to generalize the structure of proofs from the case of mathematical proof to a wider range of grounds for belief. A consideration of the full range of sources of knowledge will show that the most obvious divisions among grounds for beliefs do not parallel the logical structure of these beliefs. So if the structure of abstract proofs were to be interpreted as a hierarchical structure of grounds for belief, we would have to suppose that there was a canonical source for any belief which depended only on its logical structure, a source which was often obscured by an apparent source (in testimony, for example, or in a sort of reasoning which was not the characteristic one for beliefs with that logical structure). Something like this idea of canonical grounds for beliefs has been advanced by Dummett on the basis of general considerations regarding the nature of meaning which, for him, tell against a holistic epistemology of the sort which was assumed in the first section of this paper.[12] We will not pursue the matter in any depth here. Suffice it to say that Dummett's arguments support the ascription of a hierarchical structure to abstract proof only at the cost of holding that the range of beliefs relevant to the justification of a given belief is bounded relative to the logical complexity of the belief being justified (though it must be added that this is something he would count not as a cost but as a benefit).

The two problems with the concept of a construction meet in the interpretation of rules of inference. The operations corresponding to rules of intuitionistic logic (apart from the doubtful case of *ex falso quodlibet*) are just the operations which define the hierarchy of constructions (pairing and projection operations for the &-rules, and so on). Accordingly, these rules may be held to reflect the intrinsic nature of the constructions which form their input and output. However, specifically classical rules like indirect proof cannot be understood in this way. To give a uniform interpretation of all the rules of classical logic, it seems best to regard the operations corresponding to these rules as applying to domains which are all on the same level of abstraction. Such operations may be deemed pairing and projection operations, for example, in virtue of their algebraic properties; but, if so, they should be compared with coding and decoding operations rather than with operations which construct and dismantle mathematical objects.

If we hope to use proofs to model features of classical reasoning about empirical matters, we must look to something other than mathematical constructions as an interpretation of the abstract structure of proofs. So let us turn again to the structure which we found to lie behind the use of argument in the justification of ordinary empirical belief.

III.
THE STRUCTURE OF EPISTEMIC BIOGRAPHY

The upshot of the first section was that the function of argument is to aid in the relative identification of those collections of biographical information by which beliefs are assessed. This presupposes that the relevant fragments of epistemic biography exhibit a structure which is sufficiently rich to make the relative identification possible. But the structural features this implies are still a far cry from the domains of objects and functions exhibited by categorical and hypothetical abstract proofs, to say nothing of the operations corresponding to rules of proof. This gap will be closed only by a more detailed theory of epistemic structure. In this final section, we will consider the commitments involved in the acceptance of a theory which interprets abstract proofs by a structure claimed to be implicit in the nature of epistemic justification.

A theory of this sort may be sketched in outline as follows. We begin by positing certain epistemic objects, the *reasons* for beliefs. These are the collections of partial biographical information with respect to which beliefs are assessed. The relative identification of a reason for an argument's conclusion with respect to reasons for its premises is embodied in a function from reasons to reasons which we will call an *inference*. The structure which supports the use of argument in the justification of belief consists of domains of reasons for various sentences and inferences for various arguments together with the operation of applying an inference to reasons for its premises. When we identify reasoning as deductive, we distinguish further structure, picking out certain reasons and inferences and also further operations on reasons and inferences. It is these distinguished elements of deductive structure that serve to interpret rules of proof, and their algebraic properties are expressed in the relation of proof equivalence. Proofs are equivalent on this theory when they denote the same reason for their common conclusion and do so always—that is, for any association of reasons with their undischarged assumptions and in any structured biography.

Both the kind of information constituting epistemic biography and the structure of this information are determined by the norms which govern the assessment of belief. Let us assume that these norms take the form of general epistemic principles without attempting either to state such principles or to offer any account of their source or ground. As norms, these epistemic principles are external to the actual workings of belief formation and change but not without connection to them. For, as they concern human belief, they must reflect the character of human cognition. And since they provide maxims for our judgment, they will be reflected in the development of our beliefs to the extent that this process is consciously reasoned. Still, it is important to remember that, as norms, they are to a degree independent of human reasoning generally and, especially, of the reasoning of any one individual. So a structured biography ascribed in accordance with these principles is not simply a description of one life of reason. It is this in part, but the form of the description reflects comparison with other biographies

and subsumption under general principles. These external constraints on the character of epistemic biography not only render its structure accessible to abstract description but also determine certain features of this structure.

Reasons and inferences

Now let us see what we are committed to if we ascribe to epistemic biography a structure comparable with that of abstract proofs. We will first consider issues concerning the framework of reasons and inferences, turning later to those concerning the deductive operations defined on this framework.

The most fundamental commitment is to the articulation of epistemic biography into individual reasons. One basis we can find for this articulation is the sufficiency of partial information about a biography when an individual belief is being assessed: we need only know the aspect of the biography seen from a certain perspective. This is a consequence of the generality of epistemic principles. An assessment of a belief holds no matter how we vary the biographical details which are not relevant to this assessment, and the information which is relevant can be located in the context of different epistemic biographies. Furthermore, a single biography can comprise bodies of information which are independent in their relevance to the assessment of a given belief in the sense that each can maintain its identity in the absence of the others. In general, it is the possibility of the re-identification of bodies of information in different biographical contexts which supports the individuation of reasons. While reasons need not take the form of paradigmatic biographical fragments from which all biographies are built, we can expect information to be articulated in a way that facilitates comparison of biographies. Thus we speak of richer reasoning in terms of "more reasons" rather than "a more extensive reason" because the various individual reasons may each be independently comparable with collections of information in other biographies. A belief may be assessed with respect to any one of a number of reasons for it which are found in the believer's biography (though, to assess the *believer* fairly, we must look at all the reasons for which the belief is held and somehow weigh the good against the bad).

The comparison of biographies implicit in epistemic principles does more than support the identification of individual reasons within biographies. It fills out biographies with what we may call *virtual reasons.* These are reasons that are in some way available although not embraced. When someone—Holmes, say—does and someone else, Watson, does not base a conclusion A on evidence that both share, a virtual reason corresponding to Holmes's reasoning stands as a supplement to Watson's real biography. And this is so whether or not Watson holds the belief A for other reasons. To extend an earlier figure, reasons are like social roles and virtual reasons are like roles that go unfilled.

The status of virtual reasons (like that of unfilled roles) is an embarrassment to the theorist. They arise from comparison of biographies but are of course not part of an actual biography. Neither should they be considered part of a counterfactual biography since the changes necessary to accommodate them as real reasons could alter the character of the reasons they are to stand alongside. Virtual reasons are best thought of as the various ways that beliefs, whether held or not, can be related to the actual biography. As such, they are subject to biographical constraints, and a reason in one biography may not be matched by even a virtual reason in another. My real reasons for beliefs about the events of my parents' childhood will not be supplemented by virtual reasons corresponding to direct perceptual evidence. Nevertheless, the range of virtual reasons is wide and their presence is crucial to any uniformity in the structure of epistemic biographies.

The reasons, real and virtual, found in a biography stand in a system of relations. Most important here are the relations that enable us to explain reasons by arguments, relations that support the relative identification of reasons. We assume that such relations take the form of what we are calling inferences, i.e. total functions from reasons for the premises of an argument to reasons for its conclusion. There are two parts to this assumption—that the relations are partial functions and that they are indeed total.

In assuming that the relations supporting relative identification are partial functions, we assume that when we grasp the step from premises to conclusion in an argument we grasp a general way of specifying a reason for the conclusion relative to reasons for the premises. This reflects what was said in the first section about the role of arguments. An argument serves to indicate a reason for its conclusion, and this reason is determined by the reasons for the premises together with the step from premises to conclusion. And it is only when the step from premises to conclusion has some degree of generality that we can speak of two components combining to determine a reason for the conclusion. The nearest alternative to the use of partial functions would be to divide the step from premises to conclusion into two elements—a relation (not necessarily functional) between reasons for the premises and reasons for the conclusion and a choice of the one among those reasons for the conclusion which stand in this relation to the indicated reasons for the premises. In these terms, our employment of partial functions amounts to the assumption that the generality in the step from reasons for premises to a reason for the conclusion lies not only in the relation between reasons for premises and reasons for the conclusion but also in the choice of a particular reason for the conclusion.

Beyond assuming that an inference is a functional relation, we assume that any inference indicated by an argument is defined on each choice of reasons for the argument's premises, that it is total with respect to these premises. Only if inferences are total will the indication of the inference and the indication of the reasons for the premises be independent of

one another. The range of reasons to which a nontotal inference would apply would be narrower than that indicated by its premises since among the reasons for its premises there would be reasons on which the inference was not defined. In the presence of nontotal inferences, arguments would not suffice by themselves to mark the articulation of reasoning since the grounds for the applicability of nontotal inferences to reasons could not be indicated simply by listing premises.

The assumption that inferences are total is in part a simplifying assumption, enabling us to dispense with any representation of domains of definition beyond the type of structure provided by premises and conclusions. But this assumption also involves an ascription of further structure, for we must assume that there are a sufficient number of total inferences to make it feasible to limit our consideration to them alone. And given the assumptions we will make about the deductive structure reflected by the conditional, these two aspects of the assumption are closely intertwined. In order for even the distinguished deductive inferences corresponding to instances of *modus ponens* to be total, only total inferences can be coded by reasons for conditionals. For an inference by *modus ponens* is defined only when the inference coded by the reason for the major premise $A \quad B$ applies to the reason for the minor premise A; so the inference from $A \quad B$ and A to B by *modus ponens* can be total only if every inference coded by a reason for $A \quad B$ is total.

But these grounds for assuming totality do not support a general denial of the existence of nontotal inferences or a denial of the importance of such inferences to the structure of epistemic biography. Indeed, nontotal inferences are likely to be of interest for a wider study of this structure because of the very feature that leads us to ignore them here: their applicability is sensitive to differences among reasons for the same belief. Nor is the assumption of totality forced on us by any difficulty in generalizing the structure we describe here to partial functions. The reason for not pursuing this generalization is rather that it would contribute nothing to our understanding without an account of the features of reasons on which the application of nontotal functions depend and the way these features are indicated by argument. Any such an account is well beyond the scope of our discussion here.

A total inference identifies a reason, real or virtual, for its conclusion with respect to any choice of reasons, real or virtual, for its premises. The presence of virtual reasons is important here, for we cannot expect the real reasons of a biography to be closed under all the total inferences we ascribe to it in the course of epistemic assessment. This is so especially in the case of deductive inferences; virtual reasons fill out the biography with the results of applying a deductive inference in cases where the conclusion of the inference is not believed or is not believed for the reason that the inference identifies. Given virtual reasons, deductive inferences may be ascribed to all biographies. That is, we may regard any extension of one's grounds for belief by deductive reasoning as being always available, whether or not it is actually employed. There is some idealization here, for there is some deductive reasoning that is beyond my biological capacity as surely as is direct visual evidence of my parents' childhood. But greater realism concerning this point would leave us still with a wide range of virtual reasons derived from deductive inferences.

If the range of generality of inferences was limited to real and virtual reasons for their premises, it would be possible to state a principle of extensionality which provided conditions for the identity of inferences. Inferences could be counted as identical when they had the same premises and conclusion and identified the same reason for their conclusion with respect to any choice of reasons for their premises. But this would limit the variety of inferences associated with a given argument by the variety of reasons for its premises and conclusion, and such a limitation does not accord well with the function of inferences. A total inference serves to justify an argument considered hypothetically, as the step from its premises to its conclusion; it must show how one infers the truth of the conclusion from the truth of the premises. For this, it is not enough to indicate reasons for the conclusion only with respect to those choices of reasons for the premises that are compatible with one given biography. The generality of the inference must be tied directly to the premises themselves in a way that transcends the limited variety of reasons any one person might have for them. Consequently, we should not regard inferences as extensional with respect to real and virtual reasons.

On the other hand, in the structure of abstract proofs described in the last section, hypothetical abstract proofs were functions individuated extensionally. Although this feature of the structure of abstract proofs is not necessary for the interpretation of proofs and proof equivalence, some general laws of identity for inferences are needed. There are ways of formulating such laws without assuming extensionality for inferences with respect to real and virtual reasons or assuming any other explicit criterion of identity for inferences. Category theory provides one framework for doing this. An alternative approach is to extend the range of reasons by introducing *indeterminate* reasons and to then state an extensionality principle with respect to this wider range. This is analogous to the approach taken with abstract proofs, in which the equivalence classes *[A]/* ~ of assumptions serve as indeterminate abstract proofs and enable us to treat hypothetical abstract proofs as functions individuated extensionally. And although the use of indeterminate reasons is somewhat artificial and may seem a mere technical convenience, it is in accord with the natural idea that assuming the truth of something for the sake of argument is in some way comparable with having a reason for believing it. But there is no need to decide between the use of category theory and the use of indeterminate reasons. The important point here is the negative one: we do not assume that inferences are individuated extensionally with respect to real and virtual inferences.

Deductive structure and validity

We thus assume that all epistemic biographies share an articulation into reasons and inferences; they share a structure consisting of a collection of reasons for each sentence, a collection of inferences for each argument, and operations which apply inferences to reasons. We also assume that the domains of inferences and reasons satisfy certain closure conditions. Our assumption that inferences are total could be phrased as the assumption that these domains are closed under application. We will also assume that inferences are closed under the operation of composition for many-place functions. A third, somewhat degenerate, example of a closure condition is provided by the assumption that any argument whose conclusion is already one of its premises will have as one of the inferences associated with it in any epistemic biography the trivial *projection inference,* an inference which simply repeats the reason for this premise as a reason for the conclusion. The operations of application and composition together with the projection inferences constitute a structure that appears in all epistemic biographies as the most basic part of their deductive structure.

The inferences distinguished by this basic structure alone are all projection inferences (with identity inferences from one premise as a special case). A richer structure which distinguishes a wider range of deductive inferences appears when we consider specific logical constants. Besides distinguishing further inferences (and certain reasons) as deductive, this structure will provide further deductive operations on reasons and inferences, the introduction and elimination rules for conjunction and the conditional for example. And the structure will satisfy certain closure conditions comparable with those holding for abstract proofs. Indeed, proofs themselves provide a notation for referring to these distinguished reasons, inferences, and operations on them. However, there is no need to limit the deductive structure to logics that admit complete systems of proof. That is, we need not assume that the elements and operations constituting deductive structure may all be generated from some limited group corresponding to rules of proof. Although it is natural to assume that only logics with complete systems of proof are epistemologically significant, the theoretical framework we are considering does not force us to be dogmatic about the point.

Now, just what are we committed to in assuming that epistemic biography has a distinguished deductive structure? This assumption is of a different order from those ascribing an articulation into reasons and inferences. Those assumptions further specified the structure exhibited by the data of epistemic biography as it is used to assess justification. Although our specification of that structure has employed both idealization and representational artifice, each component has its own pedigree in epistemic principles. But we cannot expect to find such accounts for the individual elements of deductive structure piecemeal; we must look for a general rationale for the structure as a whole.

One way of finding this rationale is to consider a corresponding characterization of deductive validity; an argument is deductively valid if and only if there is some deductive inference associated with it. We are led immediately to ask for the connection between this characterization and the usual semantic one. More generally, we may ask for the connection between the deductive structure of any given biography and the semantics of the language in which beliefs are expressed. A short answer is that the connection comes by way of the epistemic roles of various elements of the language (logical particles, for example) through the expression of these roles in the general laws governing the deductive structure of all biographies. However, this invites still more questions. Although we cannot hope to settle them here, we may begin to discuss the issues raised under two headings—first, the connection between semantics and the general laws of deductive structure and, second, the connection between these general laws and their realization in the deductive structure of particular biographies.

First let us note that the distinction between deductive and nondeductive inferences has a direct epistemological motivation. This is based ultimately on the possibility of several different reasons for a given belief, a possibility which brings with it both the need to identify the reason (or reasons) for which a belief is held and the possibility of associating a variety of inferences with the same argument. Now, not all the various inferences that may be associated with a given argument need have the same epistemic status, even when the argument is valid on semantic grounds. Without considering the epistemic evaluation of inferences in detail, it is clear that one important (though not essential) feature of a good inference is that it preserve epistemic quality of reasons that is, that the reason for its conclusion that it yields on application to given reasons for its premises never be epistemically worse than the worst of the reasons for its premises.

Now, consider a trivial argument from *A* to *A*. As one of its corresponding inferences, it has the identity function on reasons for *A,* an inference which can hardly fail to preserve epistemic quality. But the identity inference is not the only one associated with an argument from *A* to *A*. For example, I may conclude "Some mammals have been dogs" from itself by the genetic principle that any dog is the offspring of two dogs and the classificatory law that all dogs are mammals. And a material inference like this cannot be expected to preserve the epistemic quality of reasons to which it is applied. The reason for the conclusion it identifies will reflect not only the reason for the premise to which it is applied but also the grounds on which I accept the genetic and classificatory principles that support the inference. Comparing this inference with an identity inference associated with the same argument, we see a general distinction among inferences which is both epistemologically significant and naturally associated with deduction.

We need not take the preservation of epistemic quality to be the defining characteristic of deductive inferences. Indeed, we can find our rationale for a distinguished deductive structure by reversing the dependence and understanding epistemic quality to be preserved by deductive inference because epistemic evaluation always respects deductive structure. That is, we should regard deductive structure as an order ascribed to epistemic biography prior to its evaluation rather than as one deriving from that evaluation. Consequently, the linguistic elements which play a role in the description of deductive structure acquire their role through ties to the ways in which structure is ascribed to the data of epistemic biography as this biography is constituted. For example, reasons for conjunctions are ascribed as codes for pairs of reasons for their conjuncts. Epistemic evaluation respects such coding, and so epistemic quality will be preserved by the inferences which exploit it (e.g. those corresponding to the &-rules). If we are to attribute deductive structure to any features of logical constants, its roots should be found in this sort of tie to the ascribed structure of epistemic biography.[13]

Part of what is needed, then, to answer our original question is an understanding of the connection between the epistemic and semantic roles of linguistic elements. And this seems inseparable from a wider understanding of the connection between epistemic norms and the semantics of the language in which belief is expressed—or, more briefly, the connection between knowledge and truth. Truth, like justification, is a necessary condition for knowledge and it might be regarded as a derivative concept complementary to justification: just as justification is what epistemic norms enable us to assess when we know only the believer's reasons, truth is what they enable us to assess when we know all but these reasons. From this perspective, semantic laws may be derived from epistemic norms. On the other hand, we might hope to characterize knowledge as securely or reliably true belief, holding epistemic norms to be derived from semantics with an eye to the exigencies of human cognition. But whether we work out such a derivation in one direction or the other, or find some more complex connection between semantics and epistemic norms, it is here that we should look for a connection between the usual semantic understanding of deduction and the general laws of deductive structure. No account of the epistemic significance of logical constants can be satisfactory except as part of an answer to these larger questions.

The second half of the original problem lies in the connection between these general laws and their particular realizations. If the first half concerned the relation of epistemology to semantics, this concerns its relation to psychology. The realization of the elements of deductive structure in particular biographies is not a purely psychological matter. An epistemic biography is ascribed as a part of the assessment of belief. It is how one's life counts in an epistemic reckoning. As a result, its structure can be as much a stipulation as a discovery. But the life that is assessed has a psychological reality, and the character of this reality affects the nature of the break between abstract laws of structure and their particular instantiations.

If we were semantically omniscient beings who differed only in contingent experience, the deductive structure of biography could be considered pure stipulation. We might suppose that epistemic norms simply directed us to construe the lives of those we evaluate in this form. But people do not always believe the deductive consequences of their beliefs. In the terms adopted here, the result of applying a deductive inference to real reasons may be only a virtual reason. And these deductive lapses can be traced neither to epistemically stipulated form nor to the material of contingent experience. In order for them to occur at definite points in an epistemic biography, there must be a tie between deductive structure and particular psychological histories. We must be able to connect deductive inferences to the processes and procedures by which beliefs have actually developed in order to locate the failures. The connection need not be simple, and there is certainly no reason to suppose that we could identify special deductive processes (or even deductive procedures outside the rare cases of explicit reasoning by rules of proof). But the connection must still exist; it is the tie between two familiar aspects of deduction, something which does no more than make explicit the implicit and yet may count as a genuine accomplishment.

This has been at most a preliminary survey of issues regarding the connection between the semantic characterization of validity and the one offered here; it does no more than indicate the terrain through which this connection runs. Its course has been left unmarked, but that should not be taken to cast doubt on either characterization. Both are natural and any obscurity in their relation should be ascribed to the concept of deduction itself. However, such obscurity may still occasion doubts about the project in which we have been engaged. There is an attitude towards these matters, finding perhaps its purest expression in the later work of Wittgenstein, according to which the appearance of obscurity at the heart of reason can be traced to attempts to discover fundamental structures in thought when any structure, as a product of thought, has only a derivative and limited significance.

Such doubts strike the account given here especially hard because more is claimed in talk of deductive structure than in talk of deductive consequence. Formally, our discussion has accorded a foundational status not only to rules of proof but also to principles of proof equivalence. While this in itself entails few commitments concerning the specific nature of these rules and principles, we have been committed to their existence. Even so, not everything that has been said above rests on the propriety of distinguishing deductive structure. One aim of this essay has been to suggest that the range of concepts employed in understanding deduction ought to be enriched in a way that reflects its epistemic role. In the specific account offered here, a universal and apodeictic character has been granted to deduction and elaborated further. If deduction has limits, the range of concepts needed to draw them can only be richer still.[14]

NOTES

1 An extended argument for a somewhat stronger claim is offered by Audi (1983).

2 Classical references for these two points are Sellars (1956) and Quine (1953), respectively.

3 A suggestion motivated by normalization is offered by Prawitz (1971: 257) and another by Stenlund (1972: ch. 5). Szabo (1978) develops the category-theoretic approach in detail; a somewhat different treatment of quantifiers is suggested by Lawvere (1969).

4 My own suggestions for the case of first-order logic with identity appear in Helman (1983).

5 These conventions can be implemented by regarding an assumption as a pair consisting of a formula and a natural number. After discharge, the numbers may be taken to indicate the points in the proof at which the discharged assumption occurs (as in a common informal notation for tree-form natural deduction proofs), but the numbers of discharged assumptions have no further significance for the identity of the proof.

6 To see the full generality of this, note that the conventions concerning discharged assumptions imply a syntactic identity between D) and for any assumption *[A]* not appearing in D.

7 An argument for the legitimacy of this definition may be appropriate. Suppose that Choose representative concrete proofs D_1 and D_2, and let *[A]* be an assumption of *A* that appears undischarged in neither. Now the same hypothetical abstract proof is associated with the contexts and so The principle of replacement in   I then tells us that the proofs are equivalent, and the remark of the preceding note tells us that they are identical, respectively, to the proofs It follows that the abstract proofs are identical.

8 There is a divergence here between the approaches based on normalization and those based on category theory. It can be extreme in the case of classical logic; see Szabo (1978: ch. 10). A way to a middle ground is suggested by Seely (1979) and another by Helman (1983).

9 Of course, we must know that the laws we have assumed to govern the structure of abstract proofs also admit a sufficient number of abstract proofs for each conclusion. That is a genuine concern since apparently innocuous assumptions about proof equivalence can have the consequence that all proofs of a given conclusion are equivalent. However, such trivialization can be shown to have been avoided only when a full set of principles of proof equivalence has been adopted, so we will not do either here. (Helman (1983) provides the needed argument for one way of filling out the present framework.)

10 This suggests that we might hope to characterize the deductive abstract proofs directly on the abstract level as those which are invariant in some sense. This problem is akin to the problem of similarly characterizing the functions definable in the typed lambda calculus. For the latter, see Plotkin (1980) and Statman (1983).

11 For a discussion and further references, see Sundholm (1983).

12 See, for example, Dummett (1978), especially pp. 312ff. Chapter 7 of Dummett (1977) contains one of a number of extended discussions of related issues.

13 This role need not be regarded as conferring a constructive meaning on the constant (the example of conjunction can be misleading in this respect). The role in deductive structure played by a compound formed using the constant need not be characterized solely by its inferential relations to logically simpler sentences.

14 I owe thanks to Emily Grosholz, Jay Garfield, and Willem de Vries for comments on earlier drafts.

REFERENCES

Audi, R. (1983) "The causal structure of indirect justification," *Journal of Philosophy* 80:398–415.

Dummett, M. (1977) *Elements of Intuitionism,* Oxford: Oxford University Press. (1978) "The justification of deduction," in Truth *and Other Enigmas,* Cambridge, MA: Harvard University Press, 290–318.

Helman, G. (1983) "An interpretation of classical proofs," *Journal of Philosophical Logic* 12:39–71.

Lawvere, F.W. (1969) "Adjointness in foundations," *Dialectica* 23:281– 96.

Plotkin, G. (1980) "Lambda-definability in the full type hierarchy," in J.P. Seldin and J.R.Hindley (eds) *To H.B.Curry: Essays on Combinatory Logic, Lambda Calculus, and Formalism,* New York: Academic Press, 363–73.

Prawitz, D. (1971) "Ideas and results in proof theory," in J.E.Fenstad (ed.) *Proceedings of the Second Scandinavian Logic Symposium,* Amsterdam: North-Holland, 235–307.

Quine, W.V.O. (1953) "Two dogmas of empiricism," in *From a Logical Point of View,* Cambridge, MA: Harvard University Press, 20–46.

Seely, R.A.G. (1979) "Weak adjointness in proof theory," in M.P. Fourman, C.J.Mulvy, and D.S.Scott (eds), *Applications of Sheaves,* Lecture Notes in Mathematics 753, Berlin: Springer-Verlag, 697–701.

Sellars, W. (1956) "Empiricism and the philosophy of mind," in H.Feigl and M.Scriven (eds) *The Foundations of Science and the Concepts of Psychology and Psychoanalysis,* Minneapolis, MN: University of Minnesota Press, 253–329.

Statman, R. (1983) " -definable functionals and   conversion," *Archiv für mathematischen Logik und Grundlagenforschung* 23:21–6.

Stenlund, S. (1972) *Combinators, Lambda Terms, and Proof Theory,* Dordrecht: Reidel.

Sundholm, G. (1983) "Constructions, proofs and the meanings of logical constants," *Journal of Philosophical Logic* 12:151–72.

3
WHAT IS A PROOF?
Richard Tieszen

SUMMARY

"What is a proof?" distinguishes several philosophical accounts of proof: formalist, platonist, empiricist, Wittgensteinian, and cognitive or phenomenological accounts. The paper then focuses on the cognitive account. On this view, a proof is what provides evidence or justification in mathematical experience, where there is no "evidence" apart from the actual or possible experience of evidence. That is, there cannot be any kind of recognition-transcendent "evidence." Following Arend Heyting, it is argued that a proof is a fulfillment of a mathematical intention, where the concept of an "intention" is to be understood in terms of a theory of intentionality. This view is considered in some detail, including the parallels drawn by Heyting, Kolmogorov, Martin-Löf and others between the proof/proposition relation and proof as the realization of an expectation, or as the solution of a problem, or as a program satisfying a particular specification. Some formal developments of this conception of proof are briefly described, as is its relation to existing model-theoretic approaches to intuitionistic logic and arithmetic.

Consider some of the philosophical answers to the question put in the title of this paper. One answer, common especially to some logicians, is that a proof is a finite configuration of signs in an inductively defined class of sign configurations of an elementary formal system. This is the formalist's conception of proof, the conception on which one can encode proofs in the natural numbers. Then for any elementary formal system in which one can do the amount of arithmetic needed to arithmetize proof, and metamathematics generally, one can prove Gödel's incompleteness theorems. Gödel's theorems are often described as showing that proof in mathematics, in the formalist's sense, is not the same thing as truth, or that syntax is not the same thing as semantics.

The formalist's conception of proof is quite alien to many working mathematicians. The working mathematician's conception of proof is not nearly so precise and well delineated. In fact, it is not so clear just what a proof is on the latter conception, but only that it is not, or not only, what a strict formalist says it is. In mathematical practice proofs involve many informal components, a kind of rigor that is independent of complete formalization, and some kind of "meaning" or semantic content. Another philosophical answer to the question, that of the (ontological) platonist, is that a proof is a mind-independent abstract object, eternal, unchanging, not located in space-time, and evidently causally inert; On this view a proof is certainly not an inductively defined syntactic object, and one might think of a proof as something not created but rather discovered by a mathematician. For an intuitionist, a proof is neither a syntactic object nor an abstract object. Rather, a proof is a mental construction, a sequence of acts that is or could be carried out in time by the mathematician. For Wittgenstein and philosophers like Dummett, a proof is a linguistic convention, a type of linguistic practice or usage. One could also distinguish empiricist, pragmatist, and other philosophical accounts of proof. On an empiricist account, for example, an inference from premises to conclusion in mathematics would be part of the fabric of our empirical knowledge. It would not be different in kind from our knowledge in the physical sciences.

What I would like to do in this paper is to discuss in some detail what I shall call the cognitive, or phenomenological, dimension of proof. The cognitive or phenomenological dimension of proof, as I shall understand it, concerns the role of proof in providing evidence or justification for a mathematical proposition, evidence that we would not possess without a proof. (Although I shall emphasize the "cognitive" aspects of proof in what follows I do not take this to imply that proof is independent of social or cultural determinations.) I mean to use the term "evidence" in such a way that there is no such thing as evidence outside the actual or possible experience of evidence. One can get some sense of the concept of "evidence" that I have in mind by reflecting on what is involved when one does not just mechanically step through a "proof with little or no understanding, but when one "sees," given a (possibly empty) set of assumptions, that a certain proposition must be true. Anyone who has written or read proofs has no doubt at one time or another experienced the phenomenon of working through a proof in such a mechanical way and knows that it is distinct from the experience in which one sees or understands a proof. My contention is that one can only be said to have evidence in mathematics in the latter case, although there are clearly differences in degrees of evidence. What is the difference between the two experiences just mentioned? The difference can apparently only reside in our *awareness* in the two cases, for nothing about the strings of symbols we write or read changes or

has to change in order for the experience to occur. To give a rough description, one might say that some form of "insight" or "realization" is involved as is, in some sense, the fact that the proof acquires "meaning" or semantic content for us upon being understood.

I would argue that a "proof" in this sense ought not to be viewed as just a syntactic object, or as a mind-independent object. It is also not something reducible to linguistic behavior along the lines of behaviorist psychology, for the sensory stimulus involved underdetermines the experience. Consider, for example, what is generally thought to be a difficult proof written in a book on number theory. No doubt people will experience this proof in quite different ways, even though it is the same stimulus that irritates their nerve endings. (And if the stimulus is not the same, or is not something that goes on at their nerve endings, then there are grave difficulties for the behaviorist about what it could be.) If this is possible then there is more to their experience than meets their sense receptors. It follows that the "proof" could not be a function of only what they receive at their nerve endings. Their responses will instead be determined by what they receive at their nerve endings plus what they actually "see" or "experience." It is necessary to make allowances, that is, for differences in the mental states or processes that are involved between stimulus and response, and it is these mental states or processes that determine whether or not one in fact has a proof. One needs an account of cognition according to which mental states can themselves be the effects of stimuli and/or other mental states, in addition to being the causes of other mental states and/or responses.

My aim in this paper then is to outline and argue for a theory of proof which does justice to what I am calling the cognitive dimension of proof. I believe that neither formalism, empiricism, pragmatism, platonism nor any of the other "isms" in the philosophy of mathematics, with the exception of intuitionism on some counts, does very well with the cognitive aspect of proof, even though it is essential for understanding mathematical knowledge.

I

A passage from some recent work of the logician Martin-Löf (1983–4:231; cf. Martin-Löf 1987:417) does a good job of capturing the cognitive or phenomenological aspect of proof that I wish to discuss. Martin-Löf says that

> the proof of a judgment is the evidence for it…thus proof is the same as evidence…the proof of a judgment is the very act of grasping, comprehending, understanding or seeing it. Thus a proof is, not an object, but an act. This is what Brouwer wanted to stress by saying that a proof is a mental construction, because what is mental, or psychic, is precisely our acts…and the act is primarily the act as it is being performed, only secondarily, and irrevocably, does it become the act that has been performed.
>
> (Martin-Löf 1983–4:231)

In the work from which this passage is drawn Martin-Löf is emphasizing the intuitionistic view that a proof is a cognitive process carried out in stages in time, a process of engaging in some mental acts in which we come to "see" or to "intuit" something. Proof, on this conception, is in the first instance an *act* or a process; only secondarily does it become an *object*. The conception of proof that Martin-Löf is describing is found very strikingly, as he notes, in Heyting's identification of proofs (or constructions) with fulfillments of intentions in the sense of Husserl's philosophy (Heyting 1931). On this view, the possession of evidence amounts to the fulfillment of mathematical intentions. I agree completely with Martin-Löf's remark (Martin-Löf 1983–4:240–1) that Heyting did not just borrow these terms from Husserl but that he also applied them appropriately. Thus, as a first approximation, I propose to answer our opening question as follows:

1 A *proof* is a fulfillment of a mathematical intention.

Likewise, provability is to be understood in terms of fulfillability. In one fell swoop this embeds the concept of proof in a rich phenomenological theory of mental acts, intentionality, evidence, and knowledge. To understand philosophically what a proof is one must understand what the fulfillment of a mathematical intention is. Let us therefore first consider the concept of a mathematical intention.

The concept of an intention is to be understood in terms of a theory of intentionality. Many cognitive scientists and philosophers of mind believe that intentionality is a basic, irreducible feature of cognition, certainly of the more theoretical forms of cognition. Intentionality is the characteristic of "aboutness" or "directedness" possessed by various kinds of mental acts. It has been formulated by saying that consciousness is always consciousness *of* something. One sees this very clearly in mathematics, for mathematical beliefs, for example, are beliefs *about* numbers, sets, functions, groups, spaces, and so on. By virtue of their "directedness" or referential character, mental acts that are intentional are supposed to be responsible for bestowing meaning, or semantic content.

A standard way to analyze the concept of intentionality is to say that acts of cognition are directed toward, or refer to objects by way of the "content" of each act, where the object of the act may or may not exist. We might picture the general structure of intentionality in the following way:

$$\text{Act(Content)} \longrightarrow \text{[object]},$$

where we "bracket" the object because we do not assume that the object of an act always exists. Phenomenologists are famous for suggesting that we "bracket" the object, and that we then focus our attention on the act (noesis) and act-content (or noema), where we think of an act as directed toward a particular object by way of its content (or noema). Whether the object exists or not depends on whether we have evidence for its existence, and such evidence would be given in further acts carried out through time.

We can capture what is essential (for our purposes) to the distinction between act and content by considering the following cases: a mathematician M might believe ϕ, that , or know ϕ, that or remember that ϕ, where ϕ is some mathematical proposition. In these cases different types of cognitive acts are involved—believing, knowing, remembering—but they have the same content, expressed by ϕ. The act-character changes, but the content is the same. Of course the content may also vary while the act-character remains the same. The content itself can have a structure that is quite complex. Also, when we say that the content is "expressed" by ϕ we shall mean that the mathematical *proposition* ϕ is an expression of the content of a particular cognitive act. Thus, there is a direct parallel between intensionality, a feature of language or expression, and intentionality, a feature of cognition, insofar as we are restricting attention to those expressions which are expressions of cognitive acts. We should not necessarily expect, for example, substitutivity *salva veritate* and existential generalization to hold for inferences involving expressions of intentions.

We could regiment our understanding of what forms ϕ can take in terms of the syntax of first-order theories. Thus when we say that an act is directed toward an object we could express this with the usual devices of first-order theories: individual constants and bound variables. For example, a mathematician might believe that *Sa* for a particular S and a particular *a,* or that (*x*)*Sx* for a particular S and a particular domain of objects D. ϕ can also have the form ()*Sx,* and so on. The restriction to first-order theories is not necessary however. In fact, there are good reasons for considering higher *types* of intentions and their fulfillments, but I shall not discuss this matter in any detail here.

The idea that an act is directed toward an object by way of its content has a direct analog in the thesis in intensional logic and mathematics that intension determines extension. We should comment on this thesis now in order to forestall some possible misunderstandings of the intentionality of cognitive acts. In particular, when we say that intension "determines" extension, or that an act is directed toward a particular object by way of its content, should we take this to mean that the act-content provides a "procedure," an algorithm, or a "process" for determining its extension? To answer yes, without qualification, would lead to a view of knowledge that is far too rigid. We shall say that the manner in which intension determines extension is one of degree, and is a function of the intention itself, background beliefs, contextual factors, and the knowledge acquired up to the present time. In the growth of knowledge intensions or concepts are themselves modified and adjusted at various stages as information and evidence is acquired. Thus, a "procedure" by which the referent of a given intention is fixed might be quite indeterminate or quite determinate, depending on how much knowledge one has acquired, or how much experience one has, with the objects in question. A belief about an object may be quite indeterminate, but even so we usually have at least some conception of how to go about improving our knowledge of the object.

We said in 1 that a proof is the *fulfillment* of a mathematical intention. One can think of mathematical intentions, or cognitive acts, as either empty (unfulfilled) or fulfilled. The difference between empty and fulfilled cognitive acts can be understood as the difference between acts in which we are merely entertaining conceptions of objects and acts in which we actually come to "see" or experience the objects of our conceptions. We shall elaborate further on the distinction in Sections III and IV. Philosophers familiar with Kant will recognize the similarity here with the Kantian distinction between conception and intuition (Kant 1973). On a Kantian view, knowledge is viewed as a product of conception and intuition. Of course this kind of distinction is not specific to Kant. It has a long history, in one form or another, in philosophy. But we might modify somewhat a famous remark of Kant, and say that in mathematics intentions (directed toward objects) without proofs are empty, but that proofs without intentions ("aboutness," and meaning or semantic content) are blind (as is perhaps the case in strict formalistic or proof-theoretic conceptions of proof). Knowledge is a product of (empty) intention and proof. In particular, the objects of acts of mathematical cognition need not exist, and we would only be warranted in asserting that they do exist if we have *evidence* for their existence. Proof is the same as evidence. Thus, as we noted above, another way of putting the distinction is to say that acts of cognition are "fulfilled" (or perhaps partially fulfilled) when we have evidence, and "unfulfilled" when we do not have evidence. Cognitive acts must be (at least partially) fulfilled if we are to have knowledge of objects. We have many beliefs and opinions in mathematics, but we only have proofs, that is, evidence, for some of these. Given our remarks about intentionality and fulfillment we should say that a necessary condition for a mathematician M to know that ϕ is that M believe ϕ, and that M's belief that ϕ be produced by a cognitive process— proof—which gives evidence for it.

To say that proof is a "process" in which an intention comes to be fulfilled is to say that it is a process of carrying out a sequence of acts in time in which we come to see an object or in which other determinations relevant to the given intention are made. This obviously places some constraints on the notions of proof and evidence, and on what can count as justification for beliefs about objects in mathematics. The process must be one a human being can carry out, on the analogy of carrying out procedures to solve problems in the empirical sciences, or else we will have unhinged the concept of evidence for objects from anything that could count as evidence for human beings, that is, from anything that humans could experience. Thus, for example, human beings can evidently not experience an infinite number of objects in a sequence of acts in a finite amount of time. One might think of mental acts, like computations, as taking place in linear time of type   . No doubt there is a sense in which we can perform classical deductions in linear time of type   , but to leave the matter there would be to miss the point. For it might be asked whether we can construct the objects these deductions may be *about* in linear time of type   , or whether we can execute operations (functions) involved in the propositions of the deduction in this time structure. The constraints on proof must be imposed all the way through the components of the propositions in a proof if we are to avoid a recognition-transcendent concept of evidence. Having said that much, however, I would argue that there are degrees and types of evidence provided by proof. We should pause over this point for a moment.

<div align="center">II</div>

We said that it is the function of a proof to provide evidence. But what kind of evidence? Intentions may be fulfilled in different ways or to different degrees. For knowledge in general we could consider the presence or absence of the following types of evidence: *a priori* evidence; evidence of "necessity"; clear and distinct evidence; intersubjective evidence; and "adequate evidence." In mathematics a fine-grained approach to questions about the adequacy of evidence is already found in proof theory, where proofs are ranked in terms of computational complexity and other measures. The classifications of proofs that result are surely relevant to questions about the "processes" which produce M's belief that ϕ. As we are viewing the matter, if one insists on a "feasible" proof, a finitist proof, an intuitionistic proof, or a predicative proof, one is insisting on a certain kind of evidence.

There are many philosophical arguments about whether proofs provide *a priori* knowledge, and whether they provide knowledge in which a conclusion follows with "necessity" from its premises. One could also discuss degrees of clarity and distinctness of proofs, and other matters like simplicity, length, as it relates to reliability, and elegance. I would like to set these questions aside here, however, since they do not pose problems specific to the conception of proof that I am discussing. Rather, I would like to comment on two issues that are especially important to the view of proof as the fulfillment of certain kinds of cognitive acts. First, one of the principal concerns about the kind of characterization of proof given in 1 has been that one then perhaps makes the concept of proof a private, subjective matter, that one flirts with solipsism. One often hears, for example, that Brouwer's conception of proof was solipsistic. Second, if there are constraints on what can count as evidence then what is the status of classical proof?

Concerning the first question, the identification of proof with the fulfillment of mathematical intentions is perfectly compatible with viewing mathematics as a social activity, and it need not entail solipsism. Arguments for the possibility of understanding cognition without falling into Cartesian difficulties go back at least as far as Kant. In spite of Brouwer's other references to Kant, it is a basic theme of Kant's philosophy that human beings are so constituted that their fundamental cognitive processes are isomorphic. This explains why there is intersubjective agreement in elementary parts of mathematics, and it shows that the investigation of cognition or intentionality need not entail commitment to personal, introspective reports, in the style of introspectionist psychology. Proof, on this view, is not a species of introspection. Recent work in cognitive science depends on a similar approach to cognition.

Thus, it could be argued that there is no such thing as a "proof" that could in principle be understood by only one person, for that would contradict the hypothesis of a universal, species-specific mental structure which makes scientific knowledge (in particular, mathematics) possible in the first place. It is not as if a proof is some kind of subjective mental content to which only one person has direct access. There have been views according to which a proof is just such a subjective mental content, but philosophers like Frege and Husserl went to great pains to refute them, correctly I believe, in their critiques of psychologism. The view of proof I would like to defend is definitely not psychologistic. Thus, as I construe 1, it is incoherent to suppose that speaking of an intention as fulfilled for a particular mathematician is not dependent on the possibility of fulfillment of the same intention for other mathematicians. Note that this does not, however, entail that there has to *in fact* be intersubjective agreement at all times and all places on all mathematical propositions. It entails only that intersubjective agreement is possible. Where *de facto* intersubjective agreement about a proof exists, knowledge is thought to have a firmer evidential foundation. This is one source of the "objectivity" of proof. The import of these remarks is perhaps best appreciated in connection with phenomena like Dedekind's famous "proof" of the existence of infinite systems, Hilbert's original solution to Gordon's problem, the history of "proofs" of Fermat's last theorem, and so on.

Concerning the second question, one might ask, for example, whether our view entails that classical proofs do not deserve to be called proofs at all. It is a delicate matter to state a position on this question that will not immediately lead to objections. Let us consider a traditional (weak) counterexample of the type that constructivists raise for the classical notion of proof. These counterexamples proceed by showing how the assumption that we have a (classical) proof of a certain kind would imply that some unsolved mathematical problem has been solved. And of course even if the particular problem chosen were to be solved, the significance of the counterexamples comes from the fact that it is possible to produce reductions to an endless number of other unsolved mathematical problems. Thus, for example, consider the decimal expansion of , and let *An* be the statement that "the nth decimal of   is a seven and is preceded by 6 sevens." Then what could it possibly mean to say that is provable? We would need to have a proof either that provides us with a natural number *n* such that *An* or that shows us that no such *n* exists. Since no such evidence is available we ought not to accept the principle of the excluded third, that is, we do not have evidence of its necessity. Or one could say that to accept it is to suppose that we have evidence or knowledge that we in fact do not possess. In some of the constructivist literature the principle is referred to, aptly, as the "principle of omniscience." In my book, *Mathematical Intuition* (Tieszen 1989), I have likened it to an "ideal of reason" in a Kantian sense, so that one could view it as an epistemically illegitimate (when applied beyond certain bounds) but unavoidable postulation of human reason, one which might nevertheless sometimes serve a purpose in human affairs.

If does not hold, we can also see that ought not to be acceptable. Thus, in a system of proof like the Gentzen-Prawitz intuitionistic system of natural deduction (e.g. Prawitz 1965), one does not have indirect proofs of the form

$$
\begin{array}{c}
\neg\phi \\
\vdots \\
\psi \wedge \neg\psi \\
\hline
\phi
\end{array}.
$$

If proof is to be defined as in 1 then, following our description of the difference between empty and fulfilled intentions, one must be directly presented with an object, or at least possess the means for becoming so presented. In an indirect proof in this form of an existentially quantified proposition one has nothing more than the contradiction obtained by assuming that an object satisfying some condition does not exist. This is not the same thing as seeing the object itself, and so cannot count as fulfillment of the intention directed to the object. Might one think of such "existence proofs" as at least providing "indirect evidence" for the existence of an object? In some contexts in mathematics it appears that this would be reasonable, because one might later find a constructive proof. Thus, in such contexts one might regard propositions proved by such means (as opposed to provably objectless propositions) as conditional or hypothetical assertions about objects for which evidence might be found at some later point in time. But there are some indirect proofs for which it appears that constructive proofs could not in principle be found. Consider, for example, the prospects for a constructive version of Cantor's indirect proof of the existence of nondenumerable totalities. Some philosophers might be tempted to adopt a form of fictionalism in the case of such "objects." However, we should perhaps not rush headlong into mathematical fictionalism in the presence of such purported proofs. Perhaps the best thing to say, in answer to our question, is not that such proofs are not part of some deductive structure, or of some systems of rules of reason, but rather that the evidence for objects simply thins out so significantly in these kinds of "proofs," as in some proofs that employ the excluded third, that one cannot hope for them to provide assurances of reliability, or consistency in our reasoning. They represent reason unbound or unconditioned, and hence are subject to the possibility of antinomies or paradoxes. Interestingly, there have been serious concerns about evidence even in the highest reaches of transfinite set theory. Thus, Gödel (Gödel 1964), for example, distinguishes the "iterative" from the "logical" conception of set and suggests that proofs in set theory about sets as objects in the iterative hierarchy do possess some degree of reliability.

III

The concept of proof described in 1 admits of substantial elaboration. Among other things, the distinction between empty and fulfilled intentions can be looked at in a number of ways. Fulfillments of intentions, in Husserl's philosophy, are also understood as realizations of expectations (Tieszen 1989). One can already see the foundations for this in ordinary perception. In ordinary perception I may have a conception of some object without actually seeing it, so that the intention directed to the object is empty. Even so, I could not help but have some expectations about the object, given the set of background beliefs I would have acquired up to that stage in time. The empty intention, that is, can be viewed as a set of anticipations or expectations about the object which are determined by background beliefs, memories, and so on, at that stage in time. Having such expectations is a fundamental feature of human cognition, especially in contexts where one is attempting to acquire knowledge. Then the expectation(s) may either be realized or not. Attempting to gain knowledge about an object, that is, is

like realizing certain expectations about the object. Of course one's expectations may have to be corrected or refined in the growth of knowledge.

The same structure is present in mathematical experience. Thus, we can think of mathematical propositions (which we are viewing as expressions of intentions) as expressions of expectations. The fulfillment of the intention is the realization of the expectation. We can therefore say that

2 A *proof* is a realization of a mathematical expectation,

where we take 2 to be true if and only if 1 is true. Speaking of a proof as a realization of an expectation, that is, should not be separated from the meaning that these terms have in a theory of intentionality. Note that while provability here is the same as "realizability," we do not mean readability in Kleene's sense. Kleene's realizability interpretations are distinct from the intended interpretation of intuitionistic logic and arithmetic, even though they do give some insight into what it means for a proof to be a realization of an expectation.

Note, incidentally, that on the view we are discussing there is a similarity in cognitive structure which cuts across perceptual and mathematical knowledge, so that there are some interesting analogies between evidence and justification in these two domains, even if some different types of evidence might be involved.

Kolmogorov's interpretation (Kolmogorov 1932) of propositions as problems or tasks, and proofs as solutions, provides another way to look at the difference between unrealized and realized expectations, or unfulfilled and fulfilled intentions. Heyting and Kolmogorov later took their interpretations to be equivalent. This appears to be perfectly sensible when one reflects on how people attempt to acquire knowledge. Attempting to acquire knowledge about an object is like solving a certain problem about the object. One solves the problem, or fails to solve it, by carrying out certain acts in time that will improve one's knowledge. This is done in a way that is evidently rule governed, for we do not just go about it randomly. The way one solves a problem has parameters which are based on the beliefs one has acquired up to the present time. One might then, by reflection, develop some insight into the rule-governed structure of the process. Thus, we can also say

3 A *proof* is the solution of a mathematical problem,

provided we take this to be true if and only if 1 is true. I do not believe it would be possible to do justice to the cognitive or phenomenological dimension of proof if one were to separate the concepts of problem and solution in 3 from a theory of intentionality.

The point about intentionality is especially important to keep in mind in the final remark I would like to make about the proof/ proposition relation. It has recently been suggested by Martin-Löf (1982) that the characterizations of proof in 1–3 are equivalent to saying that

4 A *proof* is a program which satisfies a particular specification.

That is, the procedure or method by which I fulfill an intention can be viewed as a program which satisfies a given specification. Viewing 1 and 2 in terms of 4 puts us directly in the middle of many interesting issues at the intersection of constructive mathematics, cognitive science, and artificial intelligence. Note especially that if 4 iff 1 then it would be begging some important philosophical questions in cognitive science and artificial intelligence to identify the notion of a program in 4 with that of a *machine*-computable program, e.g. Turing machine computable. If we did so, for example, it could be argued that our conception of proof is reducible to the formalist's conception that we briefly described at the beginning of the paper. It would be better, in thinking of a program as a method for fulfilling an intention, to use Errett Bishop's idea of "person programs" (Bishop 1967). Of course intuitionists have never been willing to uncritically identify human computation with machine computation, and this is reflected by the status of Church's Thesis in intuitionistic mathematics. Human programs, that is, methods for fulfilling intentions, may involve some kind of intrinsic intentionality, irreducible semantic content, consciousness, representational character, gestalt qualities, indeterminateness, implicit content, qualitative content, and so forth. Thus, provided we start with the concept of a program as a method of fulfilling an intention, 4 can also be taken as an answer to the question put in this paper.

One might ask what the implications of this view are for the kinds of "proofs" provided by automated or mechanical theorem proving. One answer would be to point out how counterintuitive it would be to suppose that machines could possess "evidence" in the sense in which we have been using this term. For that would imply that machines must be able to "experience" evidence. But machines do not experience anything or have intentionality, or at least they do not have intrinsic intentionality. One might therefore argue that automated proofs provide evidence only derivatively, only insofar as the proofs are interpreted and understood by beings with intentionality. Another answer would be to say that the problem whether machines might possess intentionality or not, or be capable of experience or not, is undecided at this stage in time. That is, that the principle of the excluded third is not provable for these propositions.

IV

It is possible to develop some of the foregoing ideas in detail in formal theories of proofs or constructions. I do not mean to suggest, however, that any particular formal theory will adequately capture the concept of proof in 1. In Martin-Löf s work the ideas are developed by way of the conception of "propositions-as-types" (e.g. Martin-Löf 1984). Martin-Löf s work is in the tradition of work by Girard, Troelstra, Howard, Lambek, Läuchli, Curry, and others. There has been a great deal of very recent research along these lines, especially in the computer science community, by de Bruijn, Reynolds, Coquand and Huet, and others. Kreisel and Goodman also did some early work on a theory of constructions, using a different approach. What these approaches have in common is the effort to directly interpret proofs as algorithmic processes (acts) or objects. In this respect they are similar to nonstandard interpretations of intuitionistic logic and arithmetic like Kleene's realizability interpretations or Gödel's *Dialectica* interpretation. The algorithmic theories can be contrasted with the semantic or model-theoretic interpretations of intuitionistic logic and arithmetic, starting with the earliest topological models up through Beth and Kripke models (see for example van Dalen and Troelstra 1988). While the model-theoretic interpretations are thought to be more or less artificial by intuitionists they none-theless capture some of the aspects of proof that we have been discussing. Let us briefly consider an example of how some of the ideas we have been discussing can be (partially) formalized in each type of approach.

In recent work on intuitionistic type theory Per Martin-Löf has developed what can be looked at as a theory of proofs which is motivated by the kind of philosophical account of proof discussed above. The concept of proof is formalized in a typed -calculus, which can be viewed as an abstract programming language. The -calculus is especially natural in this setting since it is about functions as rules, rather than as graphs, and thus represents the idea of a process of going from argument to value as coded by a definition, an idea which also preserves something of the "intensional" flavor of the notion of proof. Martin-Löf s system uses four basic forms of judgment, among which are the two that "*S* is a proposition," and "*a* is a proof (construction) of the proposition *S*." One can read these equivalently as, respectively, "*S* is an intention (expectation)," and "*a* is a method of fulfilling (realizing) the intention (expectation) *S*." Let us abbreviate this as *"a:S."* On this reading the meaning of the logical constants, for example, can be explained as follows. We use   for that which is such that there is no a such that *a*: . That is,   is a false or absurd intention like 1=2. A method of fulfilling the intention

(a) consists of *(a, b)* where *a:S* and *b: T;*
(b) consists of *i(a)* where *a:S* or *j(b)* where *b:T;*
(c) consists of *(x)b(x)* where *b(a):T* provided *a:S;*
(d) consists of *(x)b(x)* where *b(a):Sa* provided a is an individual;
(e) *(x)Sx* consists of *(a, b)* where *a* is an individual and *b:Sa;*
(f) ¬S is an abbreviation of "*S*

Martin-Löf, like Dummett (Dummett 1977), Prawitz (e.g. Prawitz 1978), and others, distinguishes canonical from noncanonical proofs and notes that this list shows methods of canonical form only. That is, it gives an explanation of what constitutes a "direct" proof of a proposition formed by means of one of the constants. The meaning of a proposition is thought to be given by what counts as a canonical proof of it (see, for example, Sundholm 1986; Prawitz 1978). In the case of conjunction, for example, a canonical proof of *S   T* consists of a proof of *S* and a proof of *T*. A noncanonical ("indirect") proof in intuitionistic mathematics is a method or program for obtaining a canonical proof. Every introduction rule, in Gentzen's sense, gives a canonical proof, while elimination rules give noncanonical proofs to their conclusions. Thus, think of the right-hand side of each clause as the premise and the left-hand side as the conclusion of an introduction rule. A canonical proof is one which has a form by which it can be directly seen that it follows from one of the rules. Martin-Löf s type theory as a whole is developed as a set of formation, introduction, elimination, and equality rules. Thus, for example, the introduction rule for natural numbers is

$$0 \in N \qquad \frac{a \in N}{a' \in N} \; .$$

10^{10} is not obtainable by the rule even though it is an element of *N*. But we know we can bring it to the form *a* for some Or consider a proposition like We know how to obtain a canonical proof. A noncanonical, and shorter, proof would be to show that by mathematical induction (which is an elimination rule in Martin-Löf s system) and then instantiate (which is also an elimination rule).

Consider how one could justify the following conditional using the clauses It might be helpful to think of how the proof of the conditional would look in a proof tree or in a Fitch-style system of natural deduction. Let . That is, a is a proof that converts any proof *(b, c)* of *S   T* into a proof *a((b, c))* of *U* We would like a method of fulfilling (realizing) the intention (expectation) *(S   (T   U))*, so let *d:S* and *e:T*. Define a construction *k* such that *k(d):T   U*. That is, *(k(d))(e):U*. We should

set , so that, using the functional abstraction operator and . The proof needed for the conditional is a construction which carries *a* into A:, that is, *ade.a((d, e))*.

As another example consider Let *b* is a proof of *Sa*   . Suppose . Then *c(a):Sa,* and hence . Thus, . So the proof needed for the conditional is .

In order to do some mathematics one can carry this approach further by developing, as Martin-Löf does, formation, introduction and elimination rules for finite sets, natural numbers, lists, and various predicates defined by transfinite induction and recursion. There are a number of variations on this kind of theory in the literature, as well as a type-theoretic interpretation of constructive Zermelo-Fraenkel set theory, and connections with other theories.

As an example of the semantic or model-theoretic approach, consider the notion of proof in 1 in the context of Kripke models. Kripke models are especially nice for providing the kinds of weak counterexamples that are the staple of intuitionism. A Kripke model is a triple where *A* is an inhabited partially ordered set (species), *D* is a mapping from *A* into a collection of inhabited sets, and *I* is a mapping defined on pairs of elements of *A* and predicate symbols, or pairs of elements of *A* and constants such that (for   ,   *A)*

(a)
(b) for *k*-ary *P*
(c)

For 0-ary predicate symbols where

$$\beta \leqslant \alpha, \qquad (I(\alpha, P) = 1 \Rightarrow I(\beta, P) = 1).$$

D(I) is the "domain function." The interpretation of a first-order language by a Kripke model is defined inductively. We suppose that the language contains constants for all elements of and suppose a is denoted by $\bar{a}$. Where only closed formulas are considered, we have the following definition:

(a)
(b)
(c)
(d)
(e) for no   ,
(f) iff for all   and for all *b*   *D( )*,
(g) iff there is an   *D( )* such that
(h) iff for all

is usually read as "   forces *S*," or "*S* is true at   ." On the basis of 1, we can read it as "*S* is fulfilled at   ." It is worthwhile to work through each of the clauses (a)-(h) to see what they mean on this interpretation. Intuitively, we think of mathematical research as progressing in stages in time. We view *S* as the expression of an intention, and the elements of *A* as *stages* in time at which we may or may not have *evidence* about such intentions. We should have a partial ordering of the stages and not, for example, a linear ordering, because at a given stage there will typically be for a mathematician *M* various possibilities about how his/her knowledge might progress. *M* might even stop attempting to gain knowledge altogether. We think of *M* as not only having evidence for *truths* at given times, but also as possibly acquiring evidence for *objects* as time progresses. The models also postulate a monotonicity condition on fulfillment: intuitively, *M* does not forget at later stages when *S* is fulfilled at some earlier stage, and once a conclusion is reached no additional information will cause it to be rejected. This represents a certain idealization of human knowledge. If the intention expressed by S at stage a is not fulfilled, then we may think of it as *empty*. To say the intention is empty does not mean that *S* is not directed for *M,* or not meaningful for *M,* but only that at   it is not fulfilled. Hence we can speak of empty and fulfilled intentions or cognitive acts at a given stage in time. Intentions that are empty at one point in time may later come to be fulfilled, or frustrated. By viewing the models as tree models in the standard way one gets a graphic representation of fulfilled and empty intentions in alternative courses of possible experience. As in our discussion of 2, intentions *S* that are empty can be under stood as expectations or anticipations. To say that   ⊩ *S* is thus to say that the expectation expressed by *S* is realized at   .

Now define *K* ⊩ *S* iff for all   ∈ *A*,   ⊩ *S;* and ⊩ *S* iff, for all *K, K* ⊩ *S*. Also, Γ ⊩ *S* iff in each *K* such that if, for all $T \in \Gamma$, K ⊩ T, then *K* ⊩ *S*. It is known that $\Gamma \vdash S$ in intuitionistic first-order logic iff Γ ⊩ *S*. For Kripke models, proof of a proposition amounts to fulfillment in all models. One might relate this to the sense of "necessity" that a proof is supposed to provide, that is, not just fulfillment in some models but fulfillment in every model, or in every model in which assumptions are fulfilled. The countermodels provided by Kripke semantics to some propositions of classical logic show that we do not have evidence for the necessity of those propositions. However, the proof that is classical. A constructive completeness proof

has not been forthcoming. One could view Beth and related models as variations on capturing some of the ideas we have been discussing.

REFERENCES

Bishop, E. (1967) *Foundations of Constructive Analysis,* New York: McGraw-Hill.

van Dalen, D. and Troelstra, A.S. (1988) *Constructivism in Mathematics,* vols I, II, Amsterdam: North-Holland.

Dummett, M. (1977) *Elements of Intuitionism,* Oxford: Oxford University Press.

Gödel, K. (1964) "What is Cantor's continuum problem?" reprinted in P. Benacerraf and H.Putnam (eds) *Philosophy of Mathematics: Selected Readings,* 2nd edn, Cambridge: Cambridge University Press, 470–85, 1983.

Heyting, A. (1931) "The intuitionist foundations of mathematics," reprinted in P.Benacerraf and H.Putnam (eds) *Philosophy of Mathematics: Selected Readings,* 2nd edn, Cambridge: Cambridge University Press, 52–61, 1983.

Kant, I. (1973) *Critique of Pure Reason,* trans, by N.K.Smith, London: Macmillan.

Kolmogorov, A.N. (1932) "Zur Deutung der Intuitionistischen Logik," *Mathematische Zeitschrift* 35:58–65.

Martin-Löf, P. (1982) "Constructive mathematics and computer programming," in H.Rose and J.Shepherdson (eds) *Logic, Methodology and Philosophy of Science,* vol. VI, Amsterdam: North-Holland, 73–118.

——(1983–4) "On the meanings of the logical constants and the justifications of the logical laws," *Atti Degli Incontri di Logica Matematica,* vol. 2, Siena: Università di Siena, 203–81.

——(1984) *Intuitionistic Type Theory,* Napoli: Bibliopolis.

——(1987) "Truth of a proposition, evidence of a judgment, validity of a proof," *Synthese 73:*407–20.

Prawitz, D. (1965) *Natural Deduction: A Proof Theoretical Study,* Stockholm: Almqvist and Wiksell.

——(1978) "Proofs and the meaning and completeness of the logical constants," in J.Hintikka *et al.* (eds) *Essays on Mathematical and Philosophical Logic,* Dordrecht: Reidel, 25–40.

Sundholm, G. (1986) "Proof theory and meaning," in D.Gabbay and F.Guenthner (eds) *Handbook of Philosophical Logic,* vol. III, Dordrecht: Kluwer/Reidel, 471–506.

Tieszen, R. (1989) *Mathematical Intuition,* Dordrecht: Kluwer/Reidel.

4
HOW TO SAY THINGS WITH FORMALISMS
David Auerbach

SUMMARY

Recent attention to "self-consistent" (Rosser-style) systems raises anew the question of the proper interpretation of the Gödel Second Incompleteness Theorem and its effect on Hilbert's Program. The traditional rendering and consequence is defended with new arguments justifying the intensional correctness of the derivability conditions.

I

The conception of formalism as uninterpreted, but interpretable, systems that was wavering into focus during the 1920s achieved clarity in Gödel's 1931 paper. The historical irony here is that while this clarity fulfilled one of Hilbert's demands for firm foundations, the Gödel results themselves have been seen as frustrating Hilbert's central epistemological desire.

Due in part to Hilbert's vagueness in formulating his demands, and in part to the subtlety of the issues involved, there remains an ongoing debate as to whether the First Incompleteness Theorem scuttled Hilbert's Program, whether only the Second did, or whether neither did. Leaving detailed exegesis of Hilbert aside, I propose to investigate some very general issues that arise in the course of such debate.

Hilbert wanted to save mathematics from Kronecker's constructivist critique, a critique made pointed by the antinomies of set theory. Hilbert, however, shared some of Kronecker's basic epistemological scruples and thus proposed, on one interpretation of his project, to salvage the nonconstructivist part of mathematics by making it meaningless but helpful.

This instrumentalist reading of Hilbert is arguable. Or rather, "meaningless" might be taken only as Fregean meaningless. It is not that the nonfinitary mathematics is true because, *inter alia,* its terms refer, but that its terms refer because it is true. The Hilbert Program then aims at making sense of this notion of truth prior to reference. This would make Hilbert what we would now call a meaning-holist. For present purposes I need not choose between these Hilberts; the common core is the claim that the theorems of mathematics get all of their acceptability from system-wide considerations and not from the one-by-one truth of the theorems or axioms. His enduring insight was that the representation of theories about infinitistic matters as finitistic objects (formalisms) gave one a finitistic handle on the nonfinitistic. Here is the standard story in brief.

Some parts of mathematics are about finite objects and finitistically establishable properties of them. The "upper bound" on this conception of finitary mathematics is the notion of the "potentially" infinite, as exemplified by the sequence of natural numbers. (It is, of course, a matter of some dispute as to how to mark precisely, in extension, this distinction.) This part of mathematics the Hilbertian treats contentually: questions of truth and belief arise and, roughly speaking, it has a Fregean semantics continuous with that of nonmathematical concrete language.

The rest of mathematics consists of ideal pseudo-statements whose only role is the efficient and secure calculation of contentual statements. (The noninstrumentalist Hilbert would put this: the rest of mathematics consists of ideal statements whose truth is explicated by their role in the efficient and secure calculation of contentual statements.) Hilbert wanted to assuage Kronecker's epistemological doubts that rendered ideal sentences problematic. But Hilbert thought that they possessed a significance insofar as they were useful in deriving real statements. This usefulness arises because a finitistic proof of a real sentence might be unfeasibly long or difficult, whereas a proof that took a shortcut through the infinite might be shorter or at least easier to find.[1] Along with this explanation of the usefulness of the ideal, Hilbert needed a certification of its safety: that is, that there is no ideal proof of a false real proposition. This is real-soundness.

Hilbert saw the then nascent conception of formalism as vital at this point. If ideal proofs could be finitistically represented as manipulations of concrete symbols and strings of symbols, then a real proof about ideal proofs becomes a feasible notion. Hilbert also thought that the combinatorial properties of concrete symbols were the real content of arithmetic; in this way the entire actual content of mathematics becomes metamathematical. All real mathematics is about symbols (and finitary systems of symbols). This answers Kronecker's complaint that the content of mathematics should be exhausted by the elementary theory of numbers.

The Hilbertian would like a finitistic proof of real-soundness; however, real-soundness of T implies consistency of T and, if T includes finitistic reasoning, T cannot prove T's consistency. So finitistic proofs of real-soundness are a forlorn hope and the epistemological goal of Hilbert's Program is unattainable.[2]

There are two related matters in this brief story that I wish to disinter. The first is the cavalier manner in which talk of mathematics is replaced by talk of formal theories—in which theorems about formalisms are used to motivate conclusions about mathematics proper. The second is the very much related issue of the sense in which consistency is expressed in a formalism. These are not only hard to get straight, but positions on these matters are often difficult to extricate from the viewpoint being argued. I will be arguing that once these buried matters are exhumed and examined, we will be able to safely rebury Hilbert's Program.

<center>II</center>

In an earlier paper,[3] I argued that there is a considerable distance between the technical result about formalisms that Gödel sketches in Theorem XI and since known (in various generalized forms) as the Gödel Second Incompleteness Theorem and its standard gloss. The standard gloss states that no sufficiently strong formal system can prove its own consistency or (as above) T cannot prove T's consistency, for Ts that extend PA. Note that the *first* theorem makes no claim about the content of its underivable sentence. What would smooth the path from the mathematical theorem to the philosophical gloss are the Positive Expressibility Thesis (PET) and the Negative Expressibility Thesis (NET):[4]

PET   The underivable (in T) sentence of the Gödel Second Incompleteness Theorem does express consistency.
NET   No derivable (in T) sentence expresses consistency.

Detlefsen (1986), in an intriguing resurrection of Hilbert's Program, argues that NET has not been established. He exploits the fact that special problems arise in looking for a satisfying account that would establish PET and NET. These problems, and technical machinery for ameliorating them, have recently become more widely known. I sketch them here, focusing on Peano arithmetic and its formalization PA.

In proving the First Incompleteness Theorem a formal predicate is constructed that numeralwise expresses *is a derivation of,* and from this predicate is constructed the famous Gödel sentence. Letting **Prf(x, y)** be an arbitrary predicate that numeralwise expresses *is a derivation of,* the Gödel sentence is where k is the Gödel number of a formula provably (in **PA**) equivalent to Many formulas numeralwise express *is a derivation of,* and any of them will suffice for the First Theorem. If we now construct as a plausible candidate for a consistency sentence, we get the problem that some such sentences are trivial theorems. The classic example uses Rosser's derivability predicate[5] but other constructions have been studied.

So either NET is false or we can find some reason to exclude Rosser-type predicates. Now we can find some *ways* to exclude them; but in the insightful and important Chapter IV of Detlefsen (1986) it is argued that these ways are not reasons.

The way that Detlefsen considers descends from the HilbertBernays derivability conditions. In particular, the formalization of what Detlefsen calls local provability completeness, LPC, plays the crucial role. LPC is simply
    LPC   If
and its formalization is[6]
    F-LPC
F-LPC is, of course, familiar in modal guise, as it appears in modal treatments of provability: Detlefsen argues that no argument for the necessity of F-LPC, as a condition on correct derivability predicates, works.

It is instructive to look at an untenable attempt at establishing NET, due to Mostowski, to see what can go wrong. Mostowski's goal is to justify F-LPC by seeing it as part of the project of representing as many truths about provability as possible.

Bew(x) is naturally viewed as the existential generalization of **Prf(x, y)** where **Prf** represents the derivability relation. Here "represents" has the purely technical meaning of "strongly represents" or "numeralwise expresses"—whenever a is a derivation of **b,** is a theorem and whenever a is not a pro**of** of **b**, is a theorem. This enforces the extensional adequacy of the representation of *is a derivation of,* but does not yield F-LPC.

But not all truths are particular, that is, numeralwise, truths. So perhaps we should demand that all truths concerning provability be represented. The first stumbling block is that Gödel theorems show the impossibility of this for precisely the case in point. So we weaken the demand to one that as many truths as possible be represented.

But for the latter proposal to make sense we have to construe it as demanding that certain truths be represented. But which? An answer that lists F-LPC is circular with respect to the project at hand, which was justifying F-LPC. (Any variation on a purely quantitative approach that subsumes F-LPC will run into the following problem: we will not be able to codify any truth about *unprovability.* Any assertion of unprovability is tantamount to consistency.)

Detlefsen is certainly right about *this* strategy (as well as those of Prawitz and Kreisel-Takeuti) of justifying the derivability conditions. Those strategies, as Detlefsen demonstrates,[7] attempt to prove too much. We can do better by attempting less and relocating disputes about F-LPC to their proper philosophical arena. I will return to this claim in Section V.

<div align="center">III</div>

Under what conditions do we legitimately ascribe meaning to the formulas of a formalization; and what do these conditions permit by way of inference concerning the subject matter of the ascribed meaning? To make this both concrete and simple: what justifies the (correct) claim that in formalized Peano arithmetic we can prove that 2 is less than 3 or that addition is commutative? In much of what follows there is a tedious dwelling on standard technical moves that in technical contexts are treated off-handedly or invisibly; here, for philosophical purposes, we linger over them.

What makes the formalism we call Peano arithmetic arithmetic is more than the formalism; we do not need something as sophisticated as the Skolem-Löwenheim theorem to tell us that uninterpreted formalisms are uninterpreted.

Formal languages permit many interpretations, and except for perhaps in some contingent historical sense, no interpretation is privileged other than by stipulation. Furthermore, for purposes that exceed mere consideration of truth conditions in an extensional language, the nonidentity of *ways of giving* the interpretation is relevant. Note that this extends to the interpretation of quantification, via the specification of the domain.[8]

We will want an account of "ways of giving" to extend naturally to allow the usual overview of the First Theorem as proceeding by constructing a formula that says that it is not derivable; and of the Second Theorem as demonstrating the nonderivability of consistency. (The First Theorem actually has very little to do with interpretation. It is best seen as a remark about formalisms, independent of their possible interpretations, and the machinery necessary to prove it is indifferent to any rich sense of interpretation. Thus *Rosser's* proof of the First Theorem does not proceed "by constructing a formula that says it is not derivable." But this is precisely the desirable expected result yielded by the intensional account being developed.) To avoid using the awkward "ways of giving" and the already otherwise employed "interpretation," I will use "reading" to talk about the richer sense of interpreting formulas sketched here.

The recasting of the language of formalized arithmetic so that we can plausibly say that a certain sentence of arithmetic(!) says that   is a quantifier, that G says that G is not provable, that ConT says that T is consistent, is not accomplished directly, by paralleling the arithmetic case.

As with arithmetic we start with the informal language, a piece of technical but natural language, with which we theorize about the syntax of formalisms. This language has terms for such entities as $\forall$,   , and predicates like *is a formula, is a derivation from,* etc. The objects that this language treats of include the symbols of **PA** and pairs and sequences of them. Following Boolos, we call the language the language of Syntax and the informal theory we couch in it is Syntax.

We do not directly specify a reading of the language of **PA** in terms of the language of Syntax. The syntactic reading of **PA** is, unimportantly, parasitic on the arithmetic interpretation.[9] We first set up a correspondence between the *objects* of **PA** and the *objects* of Syntax. This is Gödel numbering. This is used to induce a correspondence between the terms and formulas of **PA** and the names and predicates of Syntax. The initial correspondence, at the level of the primitives of Syntax, is extension respecting—a term of **PA** corresponds to a name in Syntax by denoting the Gödel number of the object denoted by that name. Definitional compounds in Syntax are then merely mimicked in **PA.** The initial coextensiveness guarantees the various useful meta-theorems about this re-reading of **PA:** sentences of **PA** are derivable in **PA** *only if* their counterparts are demonstrable in Syntax.

This correspondence makes no mention of the axiomatization of **PA;** the same old extensional interpretation of **PA** now induces a way of reading the formulas of **PA** as syntactic remarks. It is, of course, a different collection of predicates of (conservative extensions of) **PA** that will be of reading interest; typically, those whose extensions are the usual syntactic categories and which have "natural" readings into Syntax.

We abet and exploit the closeness of the correspondence by using, for certain terms and formulas of **PA,** orthographic cousins of their syntactic counterparts.[10] A rather large amount of Syntax can be mirrored in **PA;** crudely speaking, it is the Gödel Second Incompleteness Theorem that puts an upper bound on what can be mirrored.

It may strike the reader that there is a freedom here that was not present in the arithmetic interpretation of **PA.** Certainly, the strict copying into the language of **PA** of definitions and proofs in Syntax constrains the correspondence between **PA** and Syntax more than mere co-extensiveness would. But no analogous strictures were stated concerning the primitive correspondences from which the whole is built up. But there are such strictures. The term that corresponds to a name in Syntax is a *numeral* of **PA.** Without such a stricture the Gödel theorems would not be forthcoming and certain quantificational facts about provability would not even be statable.

Let **Bew(x)** be a formula of **PA** that results from a long series of definitions in **PA,** as just outlined, such that we can read **Bew (x)** as "x is derivable." We can read it like this because (i) in the standard interpretation it is true of just the (Gödel numbers of) theorems; and (ii) it was built up in **PA** by mimicking the standard definitions of formula, axiom, variable, free for, etc. in

Syntax. It is notable that condition (i) can hold without condition (ii) and that some formulas for which condition (i) but not condition (ii) hold are sufficient for a proof of G1 but not of G2.[11]

IV

I have sketched an arithmetic reading of the language of **PA** and a syntactic reading of the same language. Thus the very same formula can be read in either way. While these readings concern, so far, the *language* of **PA,** the justification of their strictures develops from the use to which we put *formalisms* (i.e. language as well as axioms and a notion of derivation).

A use to which we put formalisms is the "capturing" of facts about the intended reading. One aim is to characterize a syntactic mechanism that isolates formulas whose readings are true. (This is the successful component of Hilbert's Program.) We do this by specifying the notion of a derivation. Furthermore, we often have in mind the additional requirement that being a derivation reflect the notion of proof in a straightforward way: readings of derivations should be proofs. This is the epistemological applicability requirement:

EAR   If A is informally provable from the principles that **T**
           formalizes, then **A** is derivable in **T**; and derivations of
           A in **T** formalize proofs of **A**.

Our account of "boldfacing," via structure-sensitive readings, yields EAR, establishing PET and NET. The real work, in particular cases, is giving enough sense to the boldface mapping to support EAR. Since *is provable* is an intensional predicate the need for a structure-sensitive notion, sketched above as a "reading," should be no surprise.[12]

Built into our characterization of the reading of formulas of **PA** as arithmetic and as syntactic was a respect for logical form. Thus, if **Fa** is read as "14 is even," then $\exists$xFx is read as "something is even." When we add a notion of derivation to that language of **PA** to get **PA** proper we also obtain a useful result: there will be a derivation of $\exists$xFx from **Fa**.

We need beware a potential confusion. In giving the syntactic reading of the language of **PA** we did not use a notion of derivation; derivation merely happened to be one of the notions of the theory, Syntax, into which we were reading the formulas of **PA**. Derivation makes a second appearance, which we are now noting, as a part of **PA**, the neutral formalism. More precisely, we now consider the full-fledged notion of *is a derivation* (of **PA**), complete with specification of the usual axioms.

The usefulness of our structure-sensitive readings exceeds the modest consequence that theorems of **PA** tell us truths of Syntax; the surplus value is that the derivations of **PA** correspond to proofs in Syntax. So the appropriate start to a justification for the usual gloss of the Gödel Second Incompleteness Theorem amounts to the observation that any purported proof of CON$_{PA}$, using the machinery formalized in **PA**, would give us a derivation of **CON$_{PA}$,** which is what the purely technical result tells us cannot happen.

Here is the situation so far. Extension-respecting readings of **PA** are inadequate even for explicating the representation of arithmetic statements in **PA**. Structure-respecting readings reflect our actual practice in reading formal formulas, and the somewhat devious case of the syntactic reading of the formulas of **PA** was partially detailed. When we add the usual axiomatization in **PA** we get useful meta-theorems linking the derivability of certain formulas in **PA** with the establishment of theorems of (informal) Syntax. (Similarly, but less to the present point, we get useful meta-theorems linking the derivability of certain formulas in **PA** with the establishment of (informal) arithmetic.)

I now need to link the observations of Section II with those of Section III. One fact Syntax can prove about **Bew** is that if $\vdash_{PA}$A then $\vdash_{PA}$A Bew([A]). So **PA,** insofar as it is adequate to Syntax, ought to be able to derive **if** $\vdash_{PA}$**A then** $\vdash_{PA}$**Bew([A])**. In other words,

F-LPC

where the notation used reflects the stipulations of note 11. And, of course, F-LPC holds because the way we constructed **Bew**[13] excludes Rosser-style predicates. Thus, this justification of F-LPC does not depend on the Mostowski, Kreisel-Takeuti or Prawitz arguments that Detlefsen rightly criticizes. However, in considering arguments for F-LPC, and hence for NET, Detlefsen presents positive arguments against F-LPC; it is to those that I turn in the next section.

V

The notion of reading partially[14] developed above is the only viable candidate I know that explains the uses to which we put formalisms. As I read Detlefsen, his real quarrel is not with such a notion, nor even with constructions of **Bew** that guarantee F-LPC. I think he thinks that even if EAR is true it has no deleterious effect on the instrumentalist Hilbertian. For, quite simply, the Hilbertian is not interested in the same notion of proof that we are.

Hilbertians should grant that F-LPC is an unarguable fact about any **Bew** that really represents derivability—and further grant that this notion of derivability nicely captures the classical notion of proof. The Hilbertian's radical proposal is that our classical notion of proof needs amendment and it is that amended notion that belongs in the construction of a consistency sentence (cf. Detlefsen 1986:121ff.). But this is prior to any formalization. The derivability conditions appear because Detlefsen finds inspiration for an amended concept of proof in the Rosser predicates.

Detlefsen is careful to say that he is not, for various reasons, proposing any particular Rosser-style concept of proof as a candidate for Hilbertian proof. He does, however, regard them as good models for the right *sort* of approach to proof on an instrumentalist basis; and he defends Rosser proofs against certain objections (see Detlefsen 1986:122ff.).

Nonetheless, I think the peculiar nature of such a notion of proof is worth dwelling on, particularly since our rich notion of a reading suggests a useful heuristic for thinking about derivability predicates.

Confronted with a derivability predicate that does not satisfy F-LPC we have reason to be wary about statements made that utilize it. Consider Edna, whose set of beliefs concerning formalisms is precisely characterized by our syntactic reading of **PA**. Amongst her beliefs are beliefs about **PA**. When Edna says that if **A** is derivable in **PA** then it is derivable in **PA** that that A is derivable in **PA**, we have every reason to believe her. For, by our stipulation about the nature of Edna's beliefs, this amounts to F-LPC. If challenged, Edna can demonstrate F-LPC, having at her command the power of induction and the inductive definitions of Syntax.

Edna's situation differs from Ralph's. Ralph's set of beliefs is also partly characterized by **PA** together with a deviant syntactic reading: Ralph's beliefs about formalisms are characterized in terms of some Rosser-style derivability predicate. That is, when Ralph uses *derivable* in a belief we read it as a Rosser-style predicate. Unlike Edna, Ralph believes that CON_{PA}, although our report of this would be less misleading if we said: Ralph believes that CON^*_{PA}. Note that $\mathbf{CON_{PA}}$ and $\mathbf{CON^*_{PA}}$ are distinct, nonequivalent sentences of **PA.** In fact, an even less misleading report of Ralph's belief is:

Ralph believes that the largest consistent subsystem[15] of **PA** is consistent.

That **PA** is consistent and Ralph's belief that the largest consistent subsystem of PA is consistent are radically different beliefs, although **PA** *is* the largest consistent subsystem of **PA**. Leopold, a skeptic about PA's consistency, would hardly be reassured by Ralph's assertion of his belief, once he understood the content of Ralph's belief. Nor, of course, can Edna reassure him since she does not believe that **PA** is consistent. Of course, properly understood, neither does Ralph. Neither Edna nor Ralph have the right *de re* belief about **PA** necessary to assuage Leopold's worries.

We know that

1 Ralph can use his notion of derivability to prove anything Edna can, and
2 Ralph knows that his notion of derivability will never produce a proof of   .

But Ralph does not know 1.

The recommendation that we reform our mathematical practice and replace the canonical notion of derivability with a Rosser-style one will indeed assure us, quite easily, of consistency. But that epistemic gain is offset by the epistemic loss occasioned by not knowing what it is that is consistent.

The epistemic trade-off is most easily seen by looking at Feferman's version of a Rosser system. Taking some enumeration of the (infinitely many) axioms of **PA**, we can define the notion of an *initial* set of axioms. Then let R={x|x is a finite, initial, consistent set of axioms} be Ralph's version of the axioms of PA. R is the set of axioms of **PA**. Ralph's notion of derivation now has the following character.[16]

In reasoning about the formalism derived from R, Ralph generates ordinary derivations, d_1, d_2, d_3, …. If d_i's last line is   . he goes back and tosses out all derivations using axioms no smaller than the largest one in d_i. He proceeds, always tossing out such derivations, and every time a derivation of   is encountered the toss-out procedure is repeated. An R-derivation is a derivation that is never tossed out.

Ralph can decide whether a given derivation is an R-derivation, although neither he nor Edna can know that all derivations are Rderivations. Leopold cannot rationally worry that R-derivation is insecure. But he can, and will, be skeptical about the extent of Ralph's knowledge.

The instrumentalist can reply that this is just fine; a secure system of unknown extent is better than an insecure system of known extent. Without direct quibbling about the insecurity of **PA**, I would point out that our (nontechnical) interest in the system produced by R-derivations (or similar Rosser systems) is proportional to the strength of our belief that it is co-extensive with **PA**. But the stronger this belief, tantamount to consistency of **PA**, the less reason to bother with R-derivations.

The instrumentalist's perfectly coherent option here is to point out a divergence of interests—that "our" interest is not hers. Detlefsen's Hilbertian can live happily with a possibly pared down ideal mathematics, even one of unmappable extent. The instrumentalist is under no obligation to lay claim to our "secret" knowledge of the co-extensiveness of PA with the Rosser system. The merits (and demerits) of this choice, however, lie beyond the scope of this paper.[17]

A final point about the double role of **PA**. Unless one takes seriously the way in which the syntactic reading of **PA** yields a notion of a formalism talking about formalisms that include it, needless puzzles arise about the need for F-LPC. A useful way of looking at this involves the related "puzzle" about proofs of the Gödel Second Incompleteness Theorem for systems weaker than **PA**.

Typically, the first theorem is shown to hold for extensions of **Q** (a finitely axiomatizable system) or perhaps for some other weak system like **PRA**. The conditions on a system, needed for a proof of the Gödel First Incompleteness Theorem, are well known and simple to state. The Gödel Second Incompleteness Theorem, however, is stated in full generality for extensions of a much stronger system than **Q**, namely **PA**. Of course the *technical* reason for this is that F-LPC will not be available in weaker systems; but it is worth seeing why, conceptually, this is not a mere *ad hoc* contrivance.

What underlies the presence of F-LPC is an ability to deal with formalisms by being able to comprehend the essentially inductive definition of a formalism. Before Ralph learned as much as he did, his syntactic beliefs were characterized, not by **PA,** but by **Q**. Similarly with Edna. We maintain, for the moment, the same readings of their beliefs.

What should Leopold make of the beliefs about **PA** or **Q** entertained by the younger Ralph or Edna? They do not have any. What should we make of Edna's possible belief represented by $\mathbf{CON_Q}$? First of all she will not believe it.[18] The delicate question, however, is how we should read it or Ralph's deviant version (which he can derive). In this case neither Edna nor Ralph know what they are talking about; they do not know the first thing about formalisms. Neither of them understand (have beliefs about the defining characteristics of) formalisms in general or **PA** or **Q** in particular. They cannot even entertain the propositions about **PA**'s or **Q**'s consistency. To read these weak systems as making remarks about formalisms would be to misconstrue what they are capable of telling us.

NOTES

1 Detlefsen (1986) elaborates this in terms of human capabilities and the possible divergence between humanly natural modes of reasoning and correct modes of reasoning. This should be contrasted with the treatment of these same matters by Hallett (1989).

2 Let T be a system of ideal mathematics and let S be a finitistically acceptable theory of real mathematics. There is some question as to whether real-soundness should be taken as: [Real (A) & $\vdash_S$A] or as the weaker [Real (A) & A. On the second formulation there can be real sentences not decided by S, and if T proves them this does not obligate S to prove them. This is one way for a Hilbertian to argue the irrelevance of the First Incompleteness Theorem—the Gödel sentence, though real, is not finitistically established. Detlefsen (1990) points out that the weaker version of soundness is the plausible one.

3 Auerbach (1985).

4 From this point forward I will use the convention of boldfacing to indicate that a formal object is being referred to as well as, in the appropriate contexts, to indicate that the formal object is the formal representation of the nonbold informal term. Consider the simplest case: 2 names the numeral for 2. We shall be concerned with constraints on boldface mappings, particularly those that yield the useful **2 the only even prime**, although 2=the only even prime. Ultimately we want:

   CON(PA) is not derivable; hence CON(PA) is not provable by methods formalized by PA,

as a consequence of an adequate notion of boldfacing.

   The word "derivation" will apply to certain formal objects, while "proof refers to those unformalized items discovered in the daily practice of working mathematicians.

5 Let Prf' be

$$\mathbf{Prf(x, y)}\ \&\ \neg \exists x\ x < y\ \&\ \mathbf{Prf(x, neg(y))},$$

which reads "x is a derivation of y and there is no smaller derivation of the negation of y." For consistent formalisms like PA, **Prf'** is co-extensive with **Prf** and numeralwise expresses what **Prf** does. A more

$$\mathbf{Prf(x, y)}\ \&\ \neg\mathbf{Prf(x, [0 = 1])}.$$

The result of replacing **Prf** with either **Prf'** or **Prf''** in the "consistency" formula is a trivial theorem. This dooms numeralwise expressibility as a sufficient condition for capturing *dicta*. See Auerbach (1985) for more details.

6 See note 11.

7 See Chapter IV of Detlefsen (1986).

8 Logic texts and the technical literature are often careless about the intensional aspect of interpretation. Mates contains a brief discussion of this. Boolos, in the chapter on Peano arithmetic in the forthcoming second edition of Boolos (1978), is explicit: "S expresses the commutativity of addition because it is, as we suppose, interpreted in accordance with the usual interpretation *N* of PA, *as we standardly give that* interpretation."

9   It need not be this way. One could directly formalize Syntax in its own suitable language and prove the Gödel theorems directly for (and in) it. As far as I know it is never done quite this way. However, some approaches are certainly in this spirit; Smullyan's various abstract versions of the Gödel theorems are based on stripped down formalizations of Syntax and the detailed framework for dealing with the Gödel results in a purely syntactical manner is supplied by his *Theory of Formal Systems. Computability Theory, Semantics, and Logic Programming* by Melvin Fitting is a recent modern treatment that avoids the arithmetic route.

10   Examples help:

$$\textbf{AtForm(x)} \text{ is } \exists t < x \, \exists t' < x \, ((\textbf{Term(t)} \wedge \textbf{Term(t')} \wedge \textbf{x}$$
$$= (\ulcorner = \urcorner, t, t')) \vee x = \ulcorner \perp \urcorner)$$

where $\ulcorner = \urcorner$ and $\ulcorner \perp \urcorner$ are the numerals for the Gödel numbers of those symbols.

11   If F is a sentence of PA, how do you write in PA that F is a theorem? stripped down Rosser-style predicate is **Prf"**: Well, **F** has a Gödel number; call it f. Furthermore, there is a term in PA that names f, in fact many. We define **[F]** as the *numeral* for f. The standard way to write in PA that F is a theorem is **Bew([F])**. It is not in general true, if t is a term that denotes f, that **Bew([F])** **Bew(t)** is a theorem. If we restrict ourselves to terms for provably recursive functions then the biconditional is a theorem.

     Now suppose that F is an open sentence of PA. As it stands, both **Bew(x<y)** and **Bew([x<y])** are syntactic nonsense; we would like to give sense to such a formula so that we could say **Bew(x<y)** is true of (2, 4). Well, what do we want *this* to mean? Presumably that a certain sentence is a theorem. Not just any sentence with terms denoting 2, 4 and a formula whose extension is *is less than,* but the sentence **2<4**. So we make **Bew([F])** be true of some sequence of numbers just in case the substitution of the standard numerals for those numbers into **F** results in a theorem. Note that machinery like this is necessary even to make sense of F-LPC and to define all the appropriate varieties of term substitution.

12   Note the following feature of the boldface mapping. Let A be a sentence of English in the vocabulary of the arithmetic interpretation of **T**; and let the arithmetic formula it interprets be **A**. Suppose **A** is a theorem of **T**. Now let S be a rewriting of **A** into the vocabulary of the syntactic theory of **T**, constrained (only) by co-extcnsiveness, with the presumed Gödel numbering as the basis. Facts: Any such S is true. S need not be **A**. Indeed, S need not be a theorem. Moreover, if **A** were a non-theorem, S might be a theorem.

13   My construction of **Bew**, based on faithfulness to the notion of reading the formula syntactically, descends from the rigorous account of Feferman (1960). This assures us, modulo a concern about the representation of the axioms, of F-LPC.

     In Feferman's generalization of the Gödel Second Incompleteness Theorem, the boldface mapping (in Feferman dotting an expression corresponds to our boldface) of complex syntactic notions is achieved by straightforward transcription of their (often inductive) definitions. In particular, the derivabiliry predicate is a complex formula that encodes a usual textbook definition of *is a derivation of.* The basis of such a definition is the set of axioms. This definition of *is a derivation of* is the same across all formalisms, save for reference to the axioms. (One assumes a fixed logical apparatus.)

     How is reference to the axioms handled? Since there are, in the case of **PA**, infinitely many axioms, they are formalized via an open sentence. Many distinct open sentences will numeralwise express the same set of axioms; this creates the same state of affairs sketched above with respect to variant (and deviant) proof predicates. Only certain of the open sentences that numeralwise express the axioms of **T** *really* express the axioms of **T**. Feferman is able to characterize a property, being an "RE-formula," that guarantees correctness. The "RE" terminology comes from "recursively enumerable." In this case, it roughly means that the formula has the form of an RE definition; it *does not* mean (just) that the set picked out is recursively enumerable, but that it is picked out in a way that guarantees that the extension is recursively enumerable. Feferman (1960) and Monk (1976) give the details.

     Those who have slogged through Gödel's original paper, or some other *detailed* proof, will remember that a great deal of trouble is taken, not merely to define, in arithmetic, the right syntactic categories but to define them in certain ways: in particular, with bounds on the quantifiers. The purpose of this is to insure, not just that the sets defined are numeralwise expressible, but that the very form of the definition guarantees it. So a prover of the First Theorem shows that the definitions pick out numeralwise expressible sets by adverting to the form of the definitions. When the prover in question is **PA** itself, as in the context of the Second Theorem, we need a formalization of appropriate form. This, in effect, is what Feferman gives us with REformulas. An RE-formula is one that canonically, as a matter of form, picks out a recursively enumerable set.

     This approach individuates formalisms by their "presentation"—and co-extensive presentations are not intersubstitutable in the context of the Second Theorem.

     More precisely: if **(x)** is a formula that numeralwise expresses the axioms of T, a proof predicate can be constructed in a standard way from a. Since many as numerically define the same set of axioms, different formal proof predicates will be defined for the same axioms, one for each a. Deviant as are bizarre ways of presenting the axioms— bizarre enough to carry a trivial assurance of consistency.

14   "Partially," because I do not think enough has been said about the initial steps in the assignment; in particular the role of numerals as proper names has been left unexamined here. I leave that for another paper, where I will take up a somewhat different defense of the derivability conditions. There I will argue that Mostowski should not have aimed at all truths and settled for some, but rather should have aimed at analytic truths and gotten them all.

15   Many different notions of subsystem will do here. Ralph need understand very little about the notion of subsystem; no more, in fact, than the bare terminology suggests. For concreteness the following will do: the *n*th subsystem is the formalism characterized by the axioms <*n*. **PA** is not finitely axiomatizable, and so will have infinitely many subsystems in this sense. Some more details are supplied below.

16   Cf. Visser[1989].

17 Detlefsen, in a private communication, has pointed out to me that Detlefsen (1990) contains a discussion of this point. Detlefsen emphasizes the conceptual separability of "locative" concerns (what are the theorems) and "quality-control" concerns (soundness). Given this separability it is open to the Hilbertian to demand different sorts of evidence in the two cases. Detlefsen makes the case that the Hilbertian need not give a finitary answer to *both* concerns.

18 See Bezboruah and Shepherdson (1976).

REFERENCES

Auerbach, D. (1985) "Intensionality and the Gödel Theorems," *Philosophical Studies* 48:337–51.

Bezboruah, A. and Shepherdson, J.C. (1976) "Gödel's Second Incompleteness Theorem for **Q**," *JSL* 41:503–12.

Boolos, G. (1978) *The Unprovability of Consistency: An Essay in Modal Logic,* Cambridge: Cambridge University Press.

Detlefsen, M. (1986) *Hubert's Program: An Essay in Mathematical Instrumentalism,* Dordrecht: Reidel.

——(1990) "On an alleged refutation of Hilbert's Program using Gödel's First Incompleteness Theorem," *Journal of Philosophical Logic* 28, in press. Reprinted in this volume as Chapter 8.

Feferman, S. (1960) "Arithmetization of metamathematics in a general setting," *Fundamenta Mathematicae* 49.

Hallett, M. (1989) "Physicalism, rcductionism and Hilbert," in A.D.Irvine (ed.) *Physicalism in Mathematics,* Kluwer Academic, 182–256.

Monk, J. (1976) *Mathematical Logic,* Berlin: Springer-Verlag.

Visser, A. (1989) "Peano's smart children: a provability logical study of systems with built-in consistency," *Notre Dame Journal of Formal Logic* 30:161–96.

5
SOME CONSIDERATIONS ON ARITHMETICAL TRUTH AND THE ω-RULE

Daniel Isaacson

I dedicate this paper to my mother, Ruth E.Isaacson.

SUMMARY

This paper explores a notion of arithmetical truth with respect to which it is claimed that Peano arithmetic is complete. A proposition in a language of arithmetic is arithmetical in the sense at issue here if its truth or falsity is perceivable directly on the basis of an articulation of our grasp of the fundamental nature and structure of the natural numbers, or directly from statements which themselves are arithmetical. The claim that Peano arithmetic is complete with respect to this notion of arithmetical truth is tested by considering whether the ω-rule in some form which allows determination of the truth value of a proposition undecided by Peano arithmetic can itself be arithmetically justified.

If everything provable in a given system for arithmetic is true, then the truth condition for the universal quantifier validates extension of that system of arithmetic by one application of the ω-rule. It is argued, however, that the claim that everything provable in Peano arithmetic is true, while itself true, is not arithmetically true, and similarly for the weaker claim, sufficient to validate one application of the ω-rule to prove a true statement where , that for each natural number *n,* if then . It is then noted that the ω-rule is validated by ω-consistency, but that ω-consistency of Peano arithmetic, even restricted to primitive recursive formulas, is not arithmetically justifiable, since $_0$ transfinite induction or $_1$-reflection and equivalently the ParisHarrington statement are none of them arithmetical. We then consider extension of Peano arithmetic by the finitely applied ω-rule, in which the infinitary condition of the premise of the ω-rule is finitely expressed by a single sentence using the arithmetized proof predicate for Peano arithmetic. But extension of Peano arithmetic by this form of the ω-rule is equivalent to extension of Peano arithmetic by $_0$ transfinite induction, and so still not arithmetical. If this rule is restricted to primitive recursive formulas, its justification as an extension of PA is tantamount to establishing the consistency of PA, again not arithmetical. We then consider the extension of finitistic (primitive recursive) arithmetic by the finitely applied ω-rule, and see that this extension is equivalent to transfinite induction of order type , which is indeed arithmetically justifiable, since it is provable in Peano arithmetic, but also thereby this extension proves nothing not already provable in PA. Finally we consider some hybrid forms of the finitely applied ω-rule potentially intermediate between the two already considered and find that, insofar as they are arithmetically justifiable, they prove nothing beyond Peano arithmetic, and insofar as they extend Peano arithmetic, they are not arithmetically justifiable.

The conclusion is that this investigation reflects the stability of this notion of arithmetical truth and its coincidence with Peano arithmetic, and also thereby encourages the view that we have identified a natural type within the space of mathematical knowledge.

I.
ARITHMETICAL TRUTH

In an earlier paper (1987) I adumbrated a conception of arithmetical truth with respect to which the first-order axiom system of Peano arithmetic was to be seen as complete. Given the existence of sentences in the language of arithmetic which are not decided by Peano arithmetic, it must be, of course, that on this conception being expressible by a sentence in a first-order language of arithmetic (say with nonlogical constants 0, S, +, •, <) does not itself render a proposition arithmetical, though this condition is indeed necessary (that a proposition, to be arithmetical, must be expressible in a first-order language in which quantification is over the natural numbers reflects the minimum condition that an arithmetical proposition must be about the natural numbers).[1] The further (inductive) condition for a proposition to be arithmetical is that its truth or falsity be perceivable directly on the basis of an articulation of our grasp of the fundamental nature and structure of the natural numbers, or directly from statements which themselves are arithmetical.

Expressed in these terms, the notion of a statement in the language of arithmetic being arithmetical is *epistemic*. It has to do with the way in which we are able to perceive the statement's truth or falsity. At the same time, insofar as this concept proves cogent, its extension admits precise determination, namely as the set of sentences whose truth is decided by Peano arithmetic.

The axioms of Peano arithmetic are arithmetical in this sense, in that they are expressed in a first-order language of arithmetic, and are directly perceivable as true from our grasp of the fundamental nature and structure of the natural numbers, articulated in our understanding of the natural numbers as the smallest structure closed under a one-to-one operation of succession and containing an element itself not a successor of any element[2] (note that this categorical characterization is second-order).

A proposition expressible in the language of arithmetic is nonarithmetical when its truth, or if it is false that of its negation, cannot be seen *directly* from our grasp of the basic structure of the natural numbers, and any proof of it uses concepts whose properties required in the proof are not themselves graspable in purely arithmetical terms. Such concepts are hidden, in not occurring overtly in the expression of the proposition in the language of arithmetic but being nonetheless present, as revealed in the proof of the proposition. They are higher-order in a sense which I will discuss presently. In the case of the Gödel sentence for Peano arithmetic, the hidden concepts are provability in the formal system of Peano arithmetic and, most crucially, consistency of Peano arithmetic. That is, to perceive the truth of the Gödel sentence (presented purely in the first-order language of arithmetic) we must understand that it expresses the condition that this sentence is not provable in this given formal system *and* see that this formal system is consistent. To cite another example of incompleteness of Peano arithmetic, let me also mention Goodstein's theorem, which asserts finiteness of sequences of natural numbers generated by a given simple effective operation. Determination of the truth of this result is by the perception that Goodstein sequences of natural numbers code descending sequences of ordinals less than ε_0, along with the perception that an ε_0 linear order is a well-ordering.

I want to claim that if a proposition expressed in the language of arithmetic has a derivation in first-order logic from arithmetical propositions, then it is itself arithmetical. (This claim possibly needs qualification in respect of truths occurring as the conclusion of unsurveyably long derivations in first-order logic; I discuss this point in Section 6 of my (1987) paper.) It might be argued, against this view, that dependence on logical derivation renders a proposition non-arithmetical, in the sense being urged here, since the justification of logical deduction is in terms of the general concept of truth, rather than from our basic conception of the natural numbers. I will address this issue briefly here, and also say something more on it later, in Section III.

The point is that an account of arithmetical truth can, and indeed must, proceed from the basis that we understand the language of arithmetic. This language includes the specifically arithmetical notions of "zero," "successor," "plus," "times," and possibly others, which we understand through our grasp of the structure of the natural numbers. But it also must include such notions as "and", "or," "not," "if…then… ," "all," "some." Our grasp of the latter notions validates the principles of first-order logic. Insofar as we understand the first-order language of arithmetic, we are in possession of a basis for perceiving the correctness of the principles of first-order logic, as well as the truth of each axiom of Peano arithmetic as it applies to the structure of the natural numbers. Accordingly, each (or at least each surveyably derivable) theorem of Peano arithmetic is arithmetical.

The question then arises whether *more* than the truths of Peano arithmetic, indeed whether *every* truth expressible in the language of arithmetic, is arithmetical in the sense I am attempting to establish. A way to test the thesis that arithmetical truth coincides with the axioms of Peano arithmetic and their logical consequences is to consider means by which truths in the language of arithmetic not provable in Peano arithmetic may be established as true. In this paper I will focus on one family of cases, based on the ω-rule, which gives rise to arguments purporting either to obliterate or to redraw this claimed boundary between arithmetical and nonarithmetical.

Before turning to the main discussion, I want to cite, by way of elucidating the idea of such a boundary, some general remarks by Gödel on the techniques and principles needed to decide statements undecided in Peano arithmetic which are nonetheless finitistically meaningful in Hilbert's sense, in particular the (standardly) arithmetized consistency statement for Peano arithmetic. The issue here is whether such statements can be decided arithmetically.

Any reasoning which is finitistic in Hilbert's sense is arithmetical, in the sense at issue here. Hilbert declares (1925:376–7) that

> the only tool at our disposal in this investigation is the same as that used for the derivation of numerical equations in the construction of number theory itself,

and this passage goes on to claim that

> it is possible to obtain in a purely intuitive and finitary way, just like the truths of number theory, those insights that guarantee the reliability of the mathematical apparatus.

Hilbert underestimated the *complexity* of finitistic reasoning required for this purpose and so did not anticipate the need for *abstract* principles to organize that complexity. The point for the present discussion is that extensions of Hilbert's finitism which offer the possibility of establishing the results called for by Hilbert's Program by providing such abstract principles are not only non-finitistic, they are also non-arithmetical, in the sense at issue in this paper.

Gödel, citing Bernays, separates out as distinct elements of Hilbert's conception of real *(inhaltlich)* mathematics the constructivist strand and the finitist strand and considers that it is the demand for the second of these which must be dropped (1958: 245). The constructivity which is achieved is nonfinitistic in Hilbert's sense by its reliance on abstract rather than purely concrete principles. Gödel characterizes the abstract concepts beyond finitary mathematics needed, in particular, to prove consistency of formal number theory in the following terms (this passage is the opening paragraph of his (1958) paper, as translated and revised (1972)):

> P.Bernays has pointed out [footnote omitted] on several occasions that, in view of the fact that the consistency of a formal system cannot be proved by any deduction procedures available in the system itself, it is necessary to go beyond the framework of finitary mathematics in Hilbert's sense in order to prove the consistency of classical mathematics or even of classical number theory. Since finitary mathematics is defined [footnote: see Hilbert's explanation in 1925:170–3 (pp. 376–9)] as the mathematics of *concrete intuition,* this seems to imply that *abstract concepts* are needed for the proof of consistency of number theory. An extension of finitism by such concepts was explicitly suggested by Bernays (1935:69 (p. 271)). By abstract concepts, in this context, are meant concepts which are essentially of the second or higher level, i.e. which do not have as their content properties or relations of *concrete objects* (such as combinations of symbols), but rather of *thought structures* or *thought contents* (e.g., proofs, meaningful propositions, and so on), where in the proofs of propositions about these mental objects insights are needed which are not derived from a reflection upon the combinatorial (space-time) properties of the symbols representing them, but rather from a reflection upon the *meanings* involved.

This passage expresses the notion of higher-order I have in mind in the idea that non-arithmetical proofs of statements in the language of arithmetic are those which require higher-order concepts.[3]

Arithmetical truth does inherently involve higher-order abstract concepts in the sense Gödel speaks of here, namely (as we have already noted) those which characterize the natural numbers as the smallest structure closed under a one-to-one mapping and containing an element which does not lie in the range of that mapping. The axioms of Peano arithmetic are seen as true on the basis of this higher-order conception, and are thereby arithmetical. The limit to being arithmetical is that no *other* meanings (= abstract concepts) than those which articulate our conception of the structure of the natural numbers enter into the perception of truth. Hence any statement whose truth can be perceived on the basis of axioms from Peano arithmetic *without* appeal to any further higher-order concepts is arithmetical. The abstract principles which are constructive but not finitistic by which, for example, the consistency of Peano arithmetic can be proved (such as transfinite induction over a primitive recursive natural well-ordering of order type ϵ_0, or primitive recursion in higher types), are not only not finitistic, they are not arithmetical. The nonarithmetical nature of transfinite induction of order type ϵ_0 is discussed in Section IV. For now, we should note that there is no reason in general to expect constructivity to fall within the domain of the arithmetical. Constructivity is an ideal to which we may aspire in any (or for that matter all) areas of mathematics, for example in the theory of the real line. The concepts which are the vehicle of our understanding in a given domain of constructive mathematics will not necessarily include and certainly may not be restricted to those by which we understand the structure of the natural numbers and their arithmetic.

The axioms of Peano arithmetic are what result from interpreting the second-order variables of the categorical characterization of the structure of the natural numbers as ranging over sets of natural numbers definable in a (first-order) language of arithmetic which is strong enough to express all primitive recursive functions (e.g. with 0, S, +, • as function symbols). This is to say that the most natural way of reading off arithmetical truths from the most perspicuous higher-order characterization of the structure of the natural numbers results in the axioms of Peano arithmetic. Argument is still needed to establish that no other higher-order expression of the conception of natural number, or other way of extracting arithmetical truths from that conception, could result in perceiving the truth of a statement in the language of arithmetic lying beyond Peano arithmetic.

It is conceivable that the cogency of this characterization of arithmetical truth could be rendered compelling by results from mathematical logic, say as showing that Peano arithmetic is the weakest first-order theory obtained by restricting second-order variables in the principles of any categorical higher-order characterization of the structure of the natural numbers to properties which are definable in a first-order language strong enough to express all primitive recursive functions (though were such a result to obtain, some philosophical argument would still be called for to establish from it the coincidence of Peano arithmetic with our intended notion of arithmetical truth). We do not have a characterization result for Peano arithmetic

of the strength and generality of the one I have just suggested, though there are strong partial results in this direction by A.J.Wilkie (1981, 1987); also of great interest in this regard is recent work by Richard Kaye (1989).

This line of investigation is not the one which I pursue in this paper. Rather, I am concerned to motivate the idea that Peano arithmetic coincides with arithmetical truth by first arguing, as we have already considered, that every statement provable in Peano arithmetic is arithmetical, and then to confront specific arguments based on various forms of the ω-rule which purport to show that arithmetical truth lies beyond Peano arithmetic. Refuting these counterclaims does not thereby establish the correctness of the position they would have refuted, but I think that consideration of these arguments does provide a test of the central claim of this paper. In the process of seeing why and how these arguments fail to refute it, we understand better the content of the claim, and the claim gains in strength through being sustained in face of these would-be refutations.

II.
THE ω-RULE

The ω-rule, in its basic form, is the (infinitary) rule of derivation which licenses the inference, for any formula *"A(x)"* in the language of arithmetic:

$$\text{From} \qquad \frac{A(0) \; A(\underline{1}) \; A(\underline{2}) \; \ldots \; A(\underline{n}) \; \ldots}{\forall x A(x),} \qquad \text{To conclude}$$

where the formulas of the premise include, for each natural number *n,* the substitution instance of *"A(x)"* with the formal numeral $\underline{n}$ (underlined arabic numerals and variables will be taken to signify formal numerals for natural numbers, which for natural number *n* consists in the n-fold iteration of the successor function applied to zero; *"A(n)"* signifies the result of substituting for *"x"* in *"A(x)"* the formal numeral corresponding to the natural number *n*). The ω-rule gives rise to many reformulations and refinements. It is beyond the scope of this paper to consider them all. I will sketch certain main developments and consider in detail a central line of argument as it bears on the nature and extent of arithmetical truth.

Tarski appears to have been the first person to formulate the ω-rule (see Feferman (1986:213)), which he discussed in 1926 and presented in a lecture in 1927, though he did not publish the idea until 1933 (1933b: 294 (see also p. 279, footnote 2); 1933a: 259– 61 (see p. 260, footnote 1)). Hilbert formulated the ω-rule, in an implicitly constructive form, in 1931 (193 la: 491–2) (this seems to be the first published formulation of the rule), and used it further in that year (1931b).[4] Given the relative timing of this work of Hilbert and Gödel's discovery of incompleteness in formal systems for arithmetic (1931a), it is natural to wonder whether Hilbert's idea to use the ω-rule in his (1931a) paper was in response to the Gödel incompleteness theorems, or whether he had come to it for other reasons. Feferman considers this question carefully (1986: 208) and concludes that the available evidence offers no clear-cut answer. The ω-rule was formulated by Carnap in 1934 and 1935 (1934: Section 14, p. 38; and 1935) (the latter paper, originally intended as part of the former, was incorporated, with slight modifications, into its English translation (1937), as Sections 34a– i). In Carnap's case it is clear that he is resorting to the ω-rule directly in response to the Gödel incompleteness results (see, for example, Carnap 1937:39, first paragraph). Carnap's interest was in establishing a criterion of analyticity by which all the truths of arithmetic, and of mathematics in general, are analytic (which in light of Gödel's discovery of incompleteness could not, of course, be accomplished in terms of deduction from any usual axiom system).

The fundamental property of the ω-rule, in its basic formulation, is that when added as a rule of consequence to any system of arithmetic strong enough to decide atomic formulas (so in particular Peano arithmetic), the resulting system is complete with respect to all truths expressible in the language of arithmetic. This fact is established by a simple induction over the logical complexity of formulas in the language of arithmetic, with the base case given by the condition that the initial system decides atomic formulas, and the induction step for the universal quantifier given precisely by the ω-rule (Carnap establishes a version of this result by this argument (1937: Theorem 14.3, p. 40)).[5]

A considerable amount of research on the ω-rule has explored the interplay between constructivizations of it and completeness/ incompleteness with respect to true sentences in the language of arithmetic. The first substantial investigation of metamathematical properties of this rule seems to be that by Rosser (1937). Rosser considers a syntactically finite form of the rule, obtained by expressing within arithmetized syntax the hypothesis that all numerical substitution instances of a formula *"A (x)"* are provable in a given axiomatic system, and he obtains Gödel-type incompleteness theorems for axiomatic systems extended by finite iteration of this rule. Shoenfield (1959) shows completeness for a system with iteration of a constructivized version of the infinitary ω-rule, where at each application of the ω-rule (codes of) proofs of the numerical instances are generated by a recursive function. Feferman (1962) establishes a completeness result for transfinite iteration of the syntactically finite form of the ω-rule considered by Rosser (1937), with the iteration indexed by a path through the constructive ordinals. The result is sensitive to the path taken; Feferman and Spector (1962) show the existence of incompleteness for iterations along paths.

Not immediately relevant to present considerations is an aspect of the ω-rule which has been important in metamathematical research over the past forty years, namely that arithmetical systems with the ω-rule admit cut-elimination (while systems with ordinary induction do not). In application of this fact, the embedding of formal systems of arithmetic into semi-formal systems with the ω-rule (with constructive bounds given by assignment of recursive ordinals to the resulting proof trees) has been established as an important technique for investigating derivability in formal systems of arithmetic.[6]

Before taking up the main argument of the paper in the next section, I want to conclude this section with a few further historical remarks on the naming and provenance of the ω-rule. The designation " ω-rule," now more or less standard, seems particularly apt, given the universal acceptance of ω as the designation of the natural numbers ordered by succession (introduced by Cantor (1883:195)) and the standard usage of the related designations of " ω-consistency" and " ω-completeness." So far as I am aware, the term " ω-rule" occurs only some thirty years after the rule was first formulated, introduced in publications of Andrzej Mostowski and his co-workers (Sundholm (1983:1:3); see Grzegorczyk, Mostowski, and Ryll-Nardzewski (1958) (the actual phrase there is "rule ω," e.g. in footnote 1 and (6) on p. 190)). This usage is established in Shoenfield (1959).

The designation introduced by Tarski is "the rule of infinite induction" (see Tarski 1933b: 294; 1933a: 259). This phrase was used by Mostowski in his earlier writings (e.g. 1947:100) and also by Schütte (1951, 1960:168) (though in 1977 Schütte omits any descriptive designation and refers to this rule among the basic inferences of his system PA* just as (S2.0*)). Hilbert, in his introduction of the ω-rule (1931a) gives it no characteristic name, merely calling it "a new rule of inference" (1931a: 491–2). Later (1931b) he designates it as "Transfinite axiom and inference schemata," reflecting the fact, which he notes, that with this rule the usual rule for the universal quantifier corresponding to introduction of the universal quantifier can be subsumed, "transfinite axiom" being already his term for that rule, as in Hilbert (1931a: 490).[7] Carnap formulates the ω-rule without any descriptive designation, labeling it DC2 (1937:38), one of the two rules of Direct Consequence for his Language I.

It seems clear that Tarski and Hilbert were the first to formulate versions of the ω-rule, and that each did so independently of the other. Was Carnap's formulation of the ω-rule also arrived at independently? Carnap does not claim that it was, and the chronology of the work suggests that it was not. Carnap's personal contact with Tarski began in February 1930, when Tarski visited Vienna to give several lectures, "chiefly on metamathematics" (Carnap 1963:30). Given the issues of common interest, one would suppose that Tarski mentioned his idea of "infinite induction" in the discussions they had at that time, if not in his lectures. In November 1930 Carnap visited Warsaw for a week, giving three lectures and "had many private conversations and discussions," including with Tarski (Carnap 1963:30–1). The main idea for Logical Syntax of Language came to Carnap in January 1931, and he set out his Language I for this investigation in the spring of that year (1963:53–4). Meanwhile Hilbert had formulated the ω-rule in a lecture to the Philosophical Society of Hamburg in December 1930 and published in 1931 (1931a). The contents of this paper were described to Gödel by Bernays in a letter of 18 January 1931 (Feferman 1986:209), and later that year Gödel published (1931b) a review of Hilbert (193 la). From Carnap's close contact with Gödel in discussing these issues during this period (Carnap 1963:53), one might suppose that Gödel would have told Carnap of Hilbert's formulation of this new rule at that time. See also Carnap's acknowledgment in the foreword to *Logische Syntax der Sprache* (1934) (1937:xvi): "On the standpoint and method of syntax, I have, in particular, derived valuable suggestions from conversations with Tarski and Gödel."

Given this seeming likelihood that Carnap's (1934) use of the ω-rule occurred in the wake of knowing of Hilbert's and Tarski's formulation of the rule, the terms in which Carnap refers to Hilbert's and Tarski's formulations are puzzling. I am grateful to Saul Kripke for the thought that this puzzlement may be resolvable by supposing that, despite the contrary suggestion of circumstances, Carnap *had* formulated the ω-rule himself before he learned of either Hilbert's or Tarski's formulations.

Carnap (1934) refers to Hilbert's use of the ω-rule (193 la, b), as well as Herbrand's (1931), and also notes that "Tarski discusses Hilbert's rule ('Rule of infinite induction' (1933b: 111))—he himself had previously, in 1927, laid down a similar one" (Carnap 1937:173). He also mentions in this passage his own rule DC2, but refrains from noting that DC2 is a version of the rule previously formulated by Hilbert and Tarski, and indeed does not discuss any connection between Hilbert's and Tarski's rule and his own, either to point out similarities, except implicitly,[8] or to draw attention to differences. If Carnap had formulated his rule DC2 as a result of hearing of the idea for it in the work of Hilbert and Tarski which he cites it would be most odd, and out of character, in my impression of Carnap (from his publications, and from biographical accounts of him by others) for him not to make explicit acknowledgment of this fact here. On the other hand, if he had arrived at the idea of this rule before he heard of the earlier work of Hilbert and Tarski, he could not acknowledge them in this way, but it also seems compatible with his character that he would not wish explicitly to claim independence, as detracting from the acknowledgment he does make of their work. So we are left with the ambiguous juxtapositions of the passage cited.[9] I conclude that we may take this passage as implicitly conveying that Carnap had arrived at his formulation of the ω-rule independently of the earlier formulations by Hilbert and Tarski.

Rosser (1937) decides to call the ω-rule "Carnap's rule" (p. 129). He cites Carnap (1935: bottom of p. 173, which corresponds to 1937:106). He also observes (footnote 3) that "This same rule appeared in a paper by Hilbert (1931a: 491, lines 13– 18)," and goes on to justify his choice of appellation with the remark, "However, Hilbert did practically nothing with the rule." This ostensible justification for preferring Carnap to Hilbert in this honorific matter does not seem well based. In any case, Rosser's naming of the rule in favor of Carnap has not been followed.

Another name cited in the development of the ω-rule is that of Novikov. Shoenfield (1959:405) states that "this rule [where we require the existence of an effective method of obtaining a proof of $A(\underline{n})$ when n is given] was suggested by Novikoff," and Mostowski (1966:70) speaks of "the effective rule ω (first proposed by Novikov)." Neither of these attributions refers to publication by Novikov.

III.
CLAIM THAT THE COMPLETENESS OF THE ω-RULE OBLITERATES THE ATTEMPTED DISTINCTION BETWEEN ARITHMETICAL AND NON-ARITHMETICAL TRUTH

The basic argument to be considered in this paper runs as follows. The correctness of the ω-rule is a consequence of our fundamental conception of natural number. Accordingly, in the same way in which this conception validates each axiom and so also each theorem of Peano arithmetic, and thereby renders them arithmetical, the ω-rule is validated, and also thereby its consequences. But its consequences include all true sentences expressible in the language of arithmetic. Hence there is no distinction to be drawn between arithmetical and non-arithmetical among the true propositions expressed in the language of arithmetic.

For the argument just given to have its intended force, the ω-rule must be justified arithmetically, in the sense at issue here. There is an argument which seems to provide the required justification. Before we consider it, we need to articulate the ω-rule more clearly by giving explicit (finite) expression to its premise, avoiding the locution "and so on" signified by the use of "…" in our original formulation (as at the beginning of Section II). We may try the following. For each formula "$A(x)$" in the language of arithmetic:

$$\text{From} \quad \frac{\text{For each natural number } n, A(\underline{n})}{\forall x A(x).} \quad \text{To conclude}$$

But this cannot be right. The force of the ω-rule as a rule of inference consists in its licensing the conclusion from being able to establish each statement of the form "$A(\underline{n})$" This force would be lost if establishing the premise meant establishing that, for each natural number n, $A(\underline{n})$ since then concluding would come to nothing more than stipulating that the domain of quantification is taken to be the natural numbers.

The premise to the ω-rule must be rather that each formula of the form "$A(\underline{n})$" is *provable* in some given system or theory. Let us call the given theory T. We are particularly interested in the case when T is Peano arithmetic, but for this stage of the discussion nothing turns on which particular theory T might be. We can express the ω-rule as follows. For each formula "$A(x)$" in the language of arithmetic:

$$\text{From} \quad \frac{\text{For each natural number } n, \text{T} \vdash A(\underline{n})}{\forall x A(x).} \quad \text{To conclude}$$

This formulation is an expression of the original ω-rule, with the difference that reference to the basis on which numerical substitution instances are established is made explicit, in terms of a given system T, and also that the first formulation overtly relies upon our knowledge that the natural numbers correspond to numerals of the form $S... S0,$ generated by finite iteration of the successor function applied to 0, while in the second formulation this knowledge is left tacit, within our grasp of universal quantification over the natural numbers and their numerals.

Suppose we know that

(1) Every sentence provable in the system T is true in the structure of the natural numbers.

We can then establish the ω-rule as an extension of T, as follows. Assume the premise of the ω-rule, i.e.

(2) For each natural number n,

From (2), by (1), we have

(3) For each natural number n, "$A(\underline{n})$" is true.

The following is the clause for the universal quantifier in the inductive definition of truth, in the case where each element of the domain is denoted by a canonical term of the language (so that the induction can be directly on the truth predicate, rather than having to be done via the satisfaction relation):

(4) For every element d of the domain, "$A(\underline{d})$" is true (in the given domain) if and only if "" is true (in that domain),

where *"$\underline{d}$"* signifies the canonical term of the language which denotes the element *d* of the domain.
   Condition (3) and half of the biconditional of (4), specialized to the case of the domain of natural numbers, yields

(5) "" is true in the domain of the natural numbers.

The basic adequacy condition on any truth predicate, "… is true," is the schema (Tarski's "Convention T")

(6) *"p"* is true if and only if *p*.

Line (5) and the appropriate instance of (6) yields

(7) ,

interpreted in the domain of the natural numbers. The derivation of (7) from (2) is by the   -rule.
   We now must consider whether this derivation is arithmetical, in the sense at issue here. In the account of this notion given in Section I it was argued that arithmetical truth is closed under (firstorder) logical consequence. By this condition, each of the inference steps of the given derivation is arithmetically acceptable.
   Accordingly, the issue comes down to the arithmetical acceptability or not of our assumption (1), and the two principles from the general theory of truth, (4) and (6).
   The closure of arithmetical truth under first-order logical consequence carries with it commitment to acceptance of the general theory of truth, since the fundamental property of logical consequence is preservation of truth,[10] and, accordingly, this notion of arithmetical truth will encompass (4) (for example in validating the principle of universal generalization) and (6).[11] There is then the question of assumption (1), that everything provable in the system T is true in the structure of the natural numbers. Let us consider this assumption specifically in respect of Peano arithmetic.
   Here is an argument which purports to establish on an arithmetical basis that everything provable in Peano arithmetic is true in the structure of the natural numbers. The point is that in seeing the axioms of Peano arithmetic as true on the basis of an articulation of our grasp of the fundamental nature and structure of the natural numbers, and that derivation in first-order logic preserves truth, we see that the axioms and theorems of Peano arithmetic are true *in* the structure of the natural numbers; and seeing this fact in this way establishes it arithmetically, in the sense of this paper. If this point is accepted, then, by the preceding derivation, the   -rule is not only correct, but arithmetically correct.
   In reply, I agree (indeed have insisted) that we do see the truth in the structure of the natural numbers of *each* axiom and theorem of Peano arithmetic, and that (as already argued) this is done in such a way as to render these truths arithmetical, in the sense at issue in this paper. But what is required for discharging the assumption of the above derivation of the   -rule is to establish the *single* claim that *every* axiom and theorem of Peano arithmetic is true in the structure of the natural numbers. The latter claim is established, not from the fundamental properties of the natural numbers, but from reflection upon our understanding of those fundamental properties. Such a process of reflection is, in the sense I have been using this notion, higher-order, and insofar as it is essential to validating the   -rule, that rule is higher-order.
   Why does being higher-order in this way count against being arithmetical? After all, our understanding of the basic concept of natural number is itself inherently higher-order. I am not denying, indeed certainly consider, that we can see that *all* the axioms and theorems of Peano arithmetic are true in the structure of the natural numbers, but the point is that doing so is a step beyond our grasp of the basic structure of the natural numbers, and so violates a constraint of minimality. What is arithmetical must be expressible in the first-order language of arithmetic (as remarked in the opening paragraph of this paper) and be perceivable as true *just* from our understanding of the basic concept of natural number. In the second stage of the process of establishing the fact that all axioms and theorems of Peano arithmetic are true in the structure of the natural numbers, we proceed beyond the body of truths initially perceivable as true from our understanding of the basic concept of natural number by the *further* process of reflecting upon the fact that the axioms arrived at by this first process are, by that process, seen as true in the structure of the natural numbers. In this second process of reflection, what we are reflecting upon is not the basic concept of the natural numbers, but our initial process of reflection upon that concept.
   The attempted justification of every instance of the   -rule given above, in the form of deriving (7) from (2), proceeds from a single premise, the statement (1). I want now to consider this argument with a weaker premise, namely just enough of the

condition given by (1) as applies to a particular instance of the ω-rule, namely for a given formula *"A(x)"* such that for each natural number *n*, but (for example the Gödel formula). The required condition is the following.

(8) For every natural number x, if , then "" is true,

where signifies the formula which results from substituting the numeral corresponding to the natural number *x* for the free variable in the formula *"A(x)"* or, equivalently,

(9) For every natural number *x,* if , then

From our preceding discussion, we have agreed that for each axiom and theorem of Peano arithmetic we see its truth in the structure of the natural numbers in such a way as to render its truth arithmetical. Accordingly, for each natural number *n,* the implication,

(10 If PA,
)

is arithmetically justified (we here require the fact that, by assumption of the case being considered, for each natural number *n,* so that the truth of the consequent of each such conditional is arithmetically justified). Does (10) provide a justification of (9)? To suppose so would be to invoke the ω-rule, as applied to the predicate

$$\text{``If } PA \vdash A(\dot{x}), \text{ then } A(\dot{x}).\text{''}$$

An argument which relies upon the general correctness of the ω-rule cannot provide justification of it. Nor can this argument be part of a justification of the ω-rule by use of induction over the complexity of the predicate to which the rule is applied, since the predicate for which the ω-rule must be assumed here is more complex than, and indeed contains as a part, the predicate with respect to which application of the ω-rule is being justified. Again we see that establishing each axiom and theorem of Peano arithmetic as arithmetically true does not thereby establish the required property of Peano arithmetic which would give arithmetical justification to the ω-rule.

   The set of premises for application of the ω-rule in its infinitary form is an ω-sequence, reflecting that this rule is to do with the natural numbers and not some other structure. Is such specificity to the natural numbers sufficient to make the ω-rule characteristic of the natural numbers, and thereby arithmetically justified, despite the failure of the arguments we have considered above? It might be claimed that it is, as demonstrated by the fact that the principle of induction is a consequence of the ω-rule (if we have *A(0)* and then by n-fold application of $\forall$- elimination and *modus ponens* there is, for each *n,* a proof of *A(n)* hence, by the ω-rule, [12] The seeming force of this argument dissipates, however, with the observation that the derivation of induction from the ω-rule makes essential use of induction, within our (informal) meta-theory, in establishing, as is required for application of the ω-rule, that for each natural number *n* there exists a proof of *A(n)* the induction step of the argument being that if we have derivations of *A(n)* and of *A(n)* → *A(n+1)* then we have a derivation of *A(n+1)*). The ω-rule transmits the arithmetical principle of induction from the metatheory to the object theory, but does not itself illuminate the principle of induction by articulating a basis on which it can be seen to hold in the structure of the natural numbers.

IV.
JUSTIFICATION OF THE ω-RULE FROM ω-CONSISTENCY

The notion of ω-consistency was introduced by Gödel (1931: 173) as a soundness condition on formal axiom systems containing arithmetic, from which it follows that the negation of the sentence Gödel showed how to construct for the given system is not provable in that system. A system T with terms (numerals) for the natural numbers is said to be ω-consistent if there is no formula *"A(x)"* in the language of T for which , …, ,…and The first Gödel incompleteness theorem establishes the existence of consistent but ω-inconsistent theories in the language of arithmetic.[13]

   ω-consistency with respect to primitive recursive formulas (generally called 1-consistency; cf. Smorynski (1977:851–2)) of Peano arithmetic immediately entails soundness of the ω-rule for primitive recursive formulas as an extension of Peano arithmetic, as follows. Let *"A(x)"* be a primitive recursive formula, for which …, … . Suppose that not Then since *"¬A(x)"* is primitive recursive for *"A(x)"* primitive recursive, and Peano arithmetic is complete with respect to Σ_1-truth, PA . But in that case PA would be 1-inconsistent. Hence , which is to say that the conclusion of the ω-rule for a primitive recursive formula follows from its premises on the assumption that PA is 1-consistent.[14]

   Clearly extension of Peano arithmetic by the ω-rule for primitive recursive formulas is complete with respect to Σ_1-truth in the language of arithmetic, that is, a proper extension of Peano arithmetic. Hence if 1-consistency of Peano arithmetic can be

arithmetically justified, in the sense at issue in this paper, then the claim that Peano arithmetic is complete with respect to arithmetical truth is refuted. Before we face this question, let us first observe that the argument we have just considered straightforwardly carries on up the arithmetical hierarchy,[15] by appealing to -consistency of extensions of PA generated by successively justified forms of the -rule, as follows (so that all truths expressible in the language of arithmetic come within its scope).

Let

$$PA^0 = PA$$

$PA^{n+1}=PA^n+$one application of the -rule for $_n$-formulas.

Lemma: For each *n*, PA^n is complete with respect to n-truth.

Proof: By induction on *n*.

<u>n=0.</u> This case is the claim that PA is complete with respect to true primitive recursive sentences, which it is.

Induction step, n>0, and assume that the result holds for all *k<n*. Let be a true $_n$-sentence (so $A(x)$ is a $_{n-1}$-formula). Since is true, for each natural number *m*, $A(\underline{m})$ is true. There are two cases.

Case 1: n=1. Then $A(\underline{m})$ is primitive recursive. Since PA is complete for true primitive recursive sentences, PA $A(\underline{m})$. Hence by one application of the -rule for primitive recursive formulas,

.

Case 2: n>1. Then where $B(x, y)$ is a $_{n-2}$-formula. Since is true, there exists a natural number *k* such that $B(\underline{m}, \underline{k})$ is true. By the induction hypothesis, PA^{n-2} is complete with respect to $_{n-2}$-truths, so . Then by existential introduction in PA^{n-2}, . Clearly for any *k*, , so . Since is $_{n-1}$, the -rule for $_{n-1}$-formulas with substitution instances provable in PA^{n-1} yields , which is to say that , as required.

Corollary: For each *n*, PA^n is complete with respect to $_{n+1}$-truth.

Proof: Let $xA(x)$ be a true $_{n+1}$-sentence. Then there is some natural number *k* such that $A(\underline{k})$ is true, and since it is n, by the Lemma, Then by existential introduction in PA^n, .

Proposition: For every natural number *k*, if PA^k is -consistent, then the following form of the -rule is sound:

$$\frac{\text{for every natural number } n,\ PA^k \vdash A(\underline{n}),}{\forall x A(x),} \quad \text{where } A(x) \text{ is } \Sigma_k$$

i.e. if the premise is true, then the conclusion is true (in the structure of the natural numbers).

Proof: Let $A(x)$ be a $_k$ formula such that, for every natural number Suppose that is not true, i.e. is true. This is a $_{k+1}$-sentence. By the Corollary to the Lemma, . But this contradicts the assumed -consistency of is true.

It is immediate that, for any theory T, if T has an -model, that is, is true in a structure whose domain is the natural numbers, then T is -consistent,16 so that a way to show a theory -consistent is to show that it has an -model. Our discussion in the previous section shows, however, that there is little prospect of giving an arithmetical justification that Peano arithmetic, or any infinitely axiomatized subsystem of it, is true in the structure of the natural numbers (has an -model). While the condition of having an -model is sufficient for -consistency, it is not necessary, that is, there are -consistent theories which have no -model.[17] Hence we should now consider other means of establishing -consistency, and in particular the 1-consistency of Peano arithmetic. Can 1-consistency of Peano arithmetic be established arithmetically, that is, can it be perceived directly on the basis of an articulation of our grasp of the fundamental nature and structure of the natural numbers, or in terms of statements which are themselves arithmetical?

Using the proof-theoretic techniques of Gentzen, 1-consistency of PA can be established from $_0$ transfinite induction (see Girard 1987:419–21). So the question for us now must be, can $_0$ transfinite induction be given arithmetical justification? I hold that it cannot.

We consider the situation in which we obtain a primitive recursive relation on the natural numbers which constitutes a wellordering of order type $_0$ on the natural numbers by coding as natural numbers the notations for ordinals $< _0$ written in Cantor normal form in base . So long as we see that this is how we have obtained this primitive recursive relation, we are able to see that this relation is well-founded and hence see the correctness of the principle of induction over that relation. We depend here on our set-theoretic understanding of the notion of ordinal and in particular our proof of the Cantor normal form theorem. But these notions are not part of our understanding of the structure of the natural numbers. On the other hand, it is clear that we can obtain transfinite ordinals in purely arithmetical terms, for example the ordinal + by defining the relation that all even numbers are less than all odd numbers, with less than between numbers of the same parity as ordinary less than. Similarly, we can code ordered *n*-tuples of natural numbers by a single number and then order the codes lexicographically, to obtain an ordering of order type ^{n} which again we can see directly to be a well-ordering.

The question is then whether by further steps of this kind, of building up from below, we can arrive at apprehension of the wellfoundedness of the primitive recursive ordering of codes of the totality of ordinals less than $_0$. The answer to this

question seems to be no. Gödel (1958, 1972) offers arguments which are indicative as to why it is not possible to build up to ε_0 induction from statements and forms of argument already established as arithmetical. He is there concerned with concrete knowledge (in Hilbert's sense) rather than the notion of arithmetical understanding which I am attempting to elucidate, but, as I discussed at the end of Section I, the points he makes in this connection carry over to the present discussion.

> For the validity of the inference-scheme of recursion up to ε_0 can surely not be made open to direct inspection, as for example it can be with ω^2. More precisely, it is no longer possible to take in with the mind's eye the many different ways in which descending sequences may be structured, so that one cannot know by inspection that each such sequence must break off. In particular if one comes to know it by moving step by step from lower to higher ordinals, this will not constitute knowing it by *inspection*. It will merely be'an an abstract knowledge, gained with the help of higher-order concepts.
>
> (Gödel 1958:243)

Gödel (1972:273) amplified this point with the following observation:

> The concretely evident steps, such as ω^2, are so small that they would have to be repeated ε_0 times in order to reach ε_0.

For the issue of being arithmetical, we must also consider the possibility of justifying ε_0 induction "from above," that is, directly from our grasp of the structure of natural numbers (in the way that the axioms of Peano arithmetic are validated), rather than by building up from below, as so far discussed. It seems clear that Dedekind's characterization of the structure of the natural numbers does not do this. What is immediately obtainable from that characterization is induction with respect to the less than relation generated by the operation of succession, that is, ω induction. But ε_0 induction is not derivable from ω induction. There will be Π^1_1-statements other than the Dedekind sentence which also characterize the structure of the natural numbers and yield ε_0 induction. However, from our point of view, these are characterizations of a well-ordering of order type ε_0 (or greater) *from* which the natural number structure of order type ω can be extracted. In this clear sense, they are an expression of our grasp not of the structure of the natural numbers, but of some other more complex structure.

The two stages of the preceding discussion indicate that ε_0 induction is not arithmetical, not by the induction clause of the inductive definition of arithmetical truth, nor by the base condition.

No transfinite induction weaker than ε_0 can prove the 1-consistency of Peano arithmetic, since transfinite induction for any ordinal less than ε_0 is provable in Peano arithmetic. Nevertheless, the 1-consistency of Peano arithmetic is weaker than ε_0 induction. Measure of relative strength is given by the result of Kreisel and Levy (1968: Section 6, in particular Theorem 12) establishing that extension of Peano arithmetic by the uniform reflection principle[18] for PA is equivalent (has the same theorems as) extension of Peano arithmetic by the schema of transfinite induction with respect to a natural well-ordering of order type ε_0 for all formulas in the language of arithmetic.

The 1-consistency of Peano arithmetic is equivalent to Σ_1-local reflection (see Smorynski 1977:852). Reflection principles are successively stronger as they go up the arithmetical hierarchy. So ε_0 induction is stronger than the 1-consistency of Peano arithmetic in that it proves statements which are stronger than the 1-consistency of PA. We might look then for a more economical proof of the 1-consistency of PA, as the test of our claim concerning the boundary of arithmetical truth.

The Paris-Harrington statement is equivalent to Σ_1-reflection (and this equivalence is arithmetical, that is, provable in Peano arithmetic) (see Paris and Harrington 1977:1141) and so to the 1-consistency of Peano arithmetic. I have earlier claimed that the Paris-Harrington statement is nonarithmetical (Isaacson 1987: 162–3). Such considerations, of their very nature, cannot be decisive since one cannot survey exhaustively all *possible* proofs of a given statement. However, at least the known proofs of the Paris-Harrington statement are visibly non-arithmetical in the sense at issue here, and there is no reason to expect that any further proof may be obtained which does not depend upon nonarithmetical properties of higher-order concepts.

These considerations conclude our investigation of the possibility of giving arithmetical justification of the ω-rule (or as restricted to primitive recursive formulas) as an extension of Peano arithmetic by establishing the ω-consistency (or 1-consistency) of Peano arithmetic. Considerations of the previous section ruled out accomplishing arithmetical justification by establishing that all such formulas provable in Peano arithmetic are true in the structure of the natural numbers. In this section we considered the use of ε_0 induction, and of more economical means, but found them all to lie beyond what is arithmetically justifiable. These considerations do not exhaust the possibilities of attempting to justify *some* form of the ω-rule, and in the next two sections I turn to consider restriction of the ω-rule in terms of conditions on establishing the premise of the rule.

V.
CONSIDERATIONS ON THE FINITELY APPLIED   -RULE IN RELATION TO
ARITHMETICAL TRUTH

The motivating idea at this stage of our investigation is to formulate the   -rule so as to render it more genuinely (usably) a rule of proof. If we are to make use of the   -rule in establishing the truth of a given proposition in the language of arithmetic, we must have some perception which justifies us in accepting each one of the infinitely many premises of the form Since the theory T is axiomatic, its proof relation can be arithmetized by a (primitive recursive) predicate, "$\mathrm{Prf}_T(x,y)\ y)$", so that the premise of an application of the   -rule, for the formula *"A(x)"*, is (finitely) expressible in arithmetized syntax by the condition

$$\forall x \exists y \mathrm{Prf}_T(y, \mathrm{sub}(\ulcorner A(z) \urcorner, \mathrm{num}(x))).$$

A single effective (provable) application of the   -rule as an extension of PA can be finitely expressed, for a given formula *"A (x)"*, as the inference step

$$\frac{\mathrm{PA} \vdash \forall x \exists y \mathrm{Prf}_{\mathrm{PA}}(y, \mathrm{sub}(\ulcorner A(z) \urcorner, \mathrm{num}(x)))}{\forall x A(x).}$$

Before turning to consider the issue of arithmetical justification of such forms of the   -rule, I want to discuss a question concerning this reliance on arithmetization of syntax. Finite formulation of the   -rule by use of arithmetization of syntax in this way gives us a means of expressing precisely the condition that we have for each natural number *n* a proof of *A(n)*. But does this appeal to coding rule out the finitely expressed   -rule as arithmetical in my intended sense? After all, the proof relation becomes "hidden" by being expressed in arithmetized syntax, and containing hidden concepts seems to be characteristic of non-arithmetical truths in the language of arithmetics, as in my earlier remarks on the nonarithmetical nature of the Gödel sentence.

I consider that use of coding does not, *per se,* render a notion or a proposition non-arithmetical. With respect to the Gödel sentence, perception of its truth does indeed require recognition that, via coding of syntax in arithmetic, it expresses that *this* sentence is unprovable in the given axiom system but more to the present point, it requires perception that the axiom system is consistent, and it is this property (actually equivalent to the truth of the Gödel sentence) which cannot be perceived arithmetically (this point will be discussed further at the end of this section). In the case of the finitely expressed   -rule, the use of coding of syntax in arithmetic constitutes an inherently apt and precise way to express what we *explicitly* intend as an effectively applied   -rule. The issue as to the arithmetical justification of the effectively applied   -rule turns not on how it is expressed, but in what terms we can establish the truth of principles which validate this rule. The point may be illustrated by considering the formulation of induction as a rule of inference: "if *A*(0) is provable and is provable, then conclude The predicate "… is provable" does not belong to the language of arithmetic *per se,* but formulating the rule of induction in these terms does not render it non-arithmetical.

Carnap (1934, 1937:173) remarks that "there is nothing to prevent the practical application of such a rule" in the course of his discussion of Hilbert's rule and Tarski's infinite induction. He does not elaborate how he considers practical application to be carried out, but one supposes he must have had in mind something like the formulation given above. Carnap offers this point as a contrast to Tarski's view on the infinitary nature of the rule (Tarski 1933b: 295) (Carnap quotes part of the following passage):

> It may be remarked in passing that such a rule, on account of its "infinitistic" character, departs significantly from all rules of inference hitherto used, that it cannot easily be brought into harmony with the current view of the deductive method, and finally that the possibility of its practical application in the construction of deductive systems seems to be problematical in the highest degree.

Tarski, in a footnote at this point, observes the relevance to this problem of metamathematical considerations but seems not to be offering the point as a means of addressing the problem:

> In contrast to all the other rules of inference the rule of infinite induction is only applicable if we have succeeded in showing that all sentences of a particular infinite sequence belong to the system constructed. But since in every phase of development of the system only a finite number of sentences is "effectively" given to us, this fact can only be established by means of metamathematical considerations.

However, in his paper of 1936 Tarski clearly does consider that treating the premise of the   -rule metamathematically renders it a genuinely usable rule of inference (p. 411):

such a rule [which always leads from true sentences to true sentences, but cannot be reduced to the old rules] is the socalled rule of infinite induction according to which the sentence A can be regarded as proved provided all the sentences A_0, A_1,… A_n,… have been proved. But this rule, on account of its infinitistic nature, is in essential respects different from the old rules. It can only be applied in the construction of a theory if we have first succeeded in proving infinitely many sentences of this theory—a state of affairs which is never realized in practice. But this defect can easily be overcome by means of a certain modification of the new rule. For this purpose we consider the sentence B which asserts that all the sentences A_0, A_1,… , A_n,… are *provable* on the basis of the rules of inference hitherto used (not that they have actually been proved). We then set up the following rule: if the sentence B is proved, then the corresponding sentence A can be accepted as proved. But here it might be objected that the sentence B is not at all a sentence of the theory under construction, but belongs to the so-called metatheory (i.e. the theory *of* the theory discussed) and that in consequence a practical application of the rule in question will always require a transition from the theory to the metatheory. In order to avoid this objection we shall restrict consideration to those deductive theories in which the arithmetic of natural numbers can be developed, and observe that in every such theory all the concepts and sentences of the corresponding metatheory can be interpreted…

It appears that Hilbert, in his first formulation (193 la) of the ω-rule, had in mind establishing by metamathematical considerations that the infinitely many premises of the ω-rule hold, since he gives as a license for taking as a basis for further deduction the condition that "it has been shown that the formula *A(z)* is a correct numerical formula when *z* is any given numeral" (p. 491). The expression of this premise cannot be by the technique of arithmetization, as we have used it, since Hilbert (strangely) did not have the idea of arithmetization of syntax. But the conceptual basis of Hilbert's Program is the recognition that formal derivability is a concrete operation of the same character as computation with numbers. The condition that every numerical instance of a given formula is finitistically provable is for Hilbert a finitistically meaningful statement, for which he would then also have the expectation that, if true, it would be finitistically provable. The form we have given to the ω-rule in the present case would, it seems, be quite acceptable to Hilbert (even though he did not arrive at that formulation himself), with the one proviso that it must be required that the premise be established finitistically, for which the condition of being provable in Peano arithmetic is then clearly too weak. In the next section we will consider the arithmetized ω-rule with the condition that the premise be finitistically provable. Herbrand's (1931) formulation of the ω-rule, following Hilbert (1931a), makes, the finite expressibility and finitistic provability of the premise fully explicit (p. 624):

Let *A(x)* be a proposition without apparent variables [i.e. quantifier-free]; if it can be proved by intuitionistic procedures that this proposition, intuitionistically considered, is true for every x, then we add to the hypotheses.

It may be worth remarking that Hilbert's second formulation of the ω-rule (1931b: 121) drops the finite expression of the premises, and what he gives then is indeed the infinitary ω-rule:

If the statement *A(z)* is correct *(richtig)* whenever *z* is a numeral, then the statement holds; in this case is said to be correct.

Can the finitely applied form of the ω-rule as an extension of Peano arithmetic be given arithmetical justification? I claim that it cannot. Feferman (1962:276, Theorem 2.19(i)) establishes the equivalence of extending an arithmetical system which is either PA or a recursively enumerable extension of PA by this form of the ω-rule (Feferman's Definition 2.16(iv), p. 274) with extending that system by the closely related, but *prima facie* stronger, schema (Definition 2.16(iii)) (stated here for the case of PA)

$$\forall x \exists y \mathrm{Prf}_{PA}(y, \mathrm{sub}(\ulcorner A(z) \urcorner, \mathrm{num}(x))) \rightarrow \forall x A(x),$$

and also the seemingly yet stronger schema (Definition 2.16(v))

$$\forall x [\exists y \mathrm{Prf}_{PA}(y, \mathrm{sub}(\ulcorner A(z) \urcorner, \mathrm{num}(x))) \rightarrow A(x)].$$

These two schemata constitute versions of the uniform reflection principle for Peano arithmetic.

As cited already in our discussion in the previous section, Kreisel and Levy (1968: Section 6, in particular Theorem 12) prove that extension of Peano arithmetic by the uniform reflection principle for PA is equivalent to (has the same theorems as) extension of Peano arithmetic by the schema of transfinite induction with respect to a natural well-ordering of order type ϵ_0 for all formulas in the language of arithmetic. Hence justification of extension of Peano arithmetic by the finitely applied ω-rule is tantamount to justification of ϵ_0 transfinite induction. But in the previous section we gave arguments to the effect that ϵ_0 transfinite induction cannot be established arithmetically, which accordingly then also show that the finitely applied ω-rule as an extension of Peano arithmetic cannot be established as arithmetical.

Even though, as we have now seen, the finitely applied -rule over Peano arithmetic is not arithmetically justifiable, it remains possible that a limited form of it might be. Let us consider the case where it is restricted to primitive recursive formulas. In Section IV we considered the basic -rule with this restriction and found reasons to consider that it could not be arithmetically justified. We should begin by noting that, on the face of it, this form of the finitely applied -rule as an extension of Peano arithmetic is weaker than the correspondingly restricted basic -rule as an extension of Peano arithmetic since its premise is stronger; that is, not only that for each natural number *n,* but that this fact is itself provable in PA:

$$PA \vdash \forall x \exists y Prf_{PA}(y, sub(\ulcorner A(z)\urcorner, num(x))).$$

Despite the apparent relative weakness (below we shall establish that this apparent weakness is actual), this restricted form of the rule is still strong enough to give a proper extension to Peano arithmetic. In particular, the Gödel sentence for PA is provable from it as follows[19] (see Feferman 1962:275 for the corresponding point for Con(PA)).

Let *g* be the Gödel number of the Gödel sentence for PA, which is then equivalent to

Claim:

The claim is established by the following argument *within* PA.

(1) Assume

By arithmetized syntax, it is provable in PA that

(2) and so, from (1) and (2),

(3) Con(PA).

By arithmetization in PA of the first half of the First Incompleteness Theorem,

(4) Con(PA)

,

so, from (3) and (4),

(5) .

Then by $\forall$-elimination, from (5),

(6) .

Arithmetization of syntax in Peano arithmetic yields completeness of provability in PA with respect to true $_1$-formulas with free variables, which includes primitive recursive formulas, and this fact is itself provable in Peano arithmetic (see Smorynski 1977:844), so

(7) We have then, from (6) and (7),

(8) which is the negation of our assumption (1). We discharge that assumption by concluding its negation, and so have

(9) on no assumptions. Then from (9) by $\forall$-introduction,

(10) as required.

Here we have an example of how the ostensibly infinite set of premises of an application of the -rule can be established by finite means. Were the present case of the -rule arithmetically justifiable, the claimed completeness of Peano arithmetic with respect to arithmetical truth would be refuted by this derivation of the Gödel sentence for PA.

While this form of the -rule properly extends Peano arithmetic, it is still relatively weak (as earlier we noted *prima facie* grounds for expecting it to be), in particular weaker than the basic -rule restricted to primitive recursive formulas (which as an extension of Peano arithmetic is complete with respect to Π_1^- truth). Here is an example of a true Π_1^- sentence not provable in PA+finitely applied -rule for primitive recursive formulas.[20] Let as a schema, with *A(x)* restricted to primitive recursive predicates. Clearly T is an axiomatic extension of PA and so has a primitive recursive proof predicate $Prf_T(u, v)$. Suppose, . Then by pure predicate logic, . This formula expresses the consistency of T which, since T is consistent, cannot, by Gödel's Second Incompleteness Theorem, be provable in T. So the supposition is false, that is

$$PA \nvdash \forall x \exists y Prf_{PA}(y, sub(\ulcorner \neg Prf_T(z, \ulcorner 0 = 1 \urcorner)\urcorner, num(x))).$$

We now face the question whether the finitely applied -rule restricted to primitive recursive formulas can be given arithmetical justification. The key to answering this question is that justification of this rule is equivalent to establishing the consistency of Peano arithmetic. We see this fact as follows.

Feferman (1962) establishes the equivalence of extending PA by the finitely applied -rule with extending PA by uniform reflection (p. 276, Theorem 2.19(i)) (which we have made use of earlier). Feferman's proof immediately yields that extension of PA by the finitely applied -rule restricted to primitive recursive formulas is equivalent to the extension of PA by uniform reflection over PA for primitive recursive formulas. But this extension proves the consistency of PA, since it includes as an instance

$$Prov_{PA}(\ulcorner 0 = 1 \urcorner) \to 0 = 1,$$

which is equivalent to , that is, Con(PA). Indeed this extension is equivalent to Con(PA), since uniform reflection over Peano arithmetic for $_1$-formulas is provable from Con(PA) (see Smorynski 1977:846, Theorem 4.1.4) and includes uniform reflection over Peano arithmetic for primitive recursive formulas as a special case. Hence justification for extending Peano

arithmetic by the finitely applied ω-rule restricted to primitive recursive formulas is tantamount to establishing the consistency of Peano arithmetic. The question we then face is whether the consistency of Peano arithmetic can be established arithmetically, in the sense at issue here.

The property of consistency is weaker than that of 1-consistency, for which we have already considered, and answered in the negative, the question of arithmetical justification. Though consistency is a weaker property, the available means by which we may establish it are closely similar to those for establishing 1-consistency. Proof theoretically, there are techniques based on appeal to transfinite induction of order type ϵ_0. We have earlier found reasons to consider that transfinite induction of order type ϵ_0 cannot be given arithmetical justification. Model theoretically, consistency of a theory follows from the existence of a model for that theory (and conversely, by the completeness theorem for firstorder logic). We can see that Peano arithmetic has a model, namely the structure of the natural numbers. But I have argued in Section III that this perception is a result of reflecting upon the process by which each axiom of Peano arithmetic is perceived as arithmetically true and that such higher-order reflection is nonarithmetical.

In its basic form, the ω-rule is validated by our understanding of the truth condition for the universal quantifier in the particular case of the structure of the natural numbers, conjoined with the perception that all the numerical instances of the premise of an instance of the ω-rule, by being provable, are true. The condition of consistency of PA which validates the extension of PA by the finitely applied ω-rule restricted to primitive recursive formulas is weaker than that soundness condition on provability in Peano arithmetic. At the same time, it is strong in relation to what can be achieved with this form of the rule. Consider the example of the proof of the Gödel sentence for PA by use of this form of the rule. The Gödel sentence for PA is equivalent to the arithmetized consistency statement for PA (with respect to the natural arithmetized proof predicate). Hence it cannot be claimed that the proof of the Gödel sentence by use of this case of the ω-rule gives us an understanding of the basis of its truth which previously we lacked. To the extent that we have a justification for applying this rule, we already have a basis for seeing the truth of that sentence.

These considerations highlight the fact that the ω-rule, while ostensibly a means by which to establish truths beyond a given theory, requires for its justification perceptions which already justify those further truths. On this basis one can see that the ω-rule, in the forms so far considered, does not extend the domain of arithmetical truths beyond Peano arithmetic.

VI.
(ALMOST) FINITISTIC APPLICATION OF THE ω-RULE IS ARITHMETICAL, BUT IT DOES NOT EXTEND PEANO ARITHMETIC

There is one further facet of the relationship between the ω-rule and arithmetical truth which I want yet to consider in this paper, namely the situation when the ω-rule (finitely expressed, as discussed in the previous section) is used to extend finitistic arithmetic.[21]

The issues which were the initial motivation for investigation of the finitistically applied ω-rule are not directly those being addressed here. The original question was whether, in face of incompleteness revealed by Gödel's results, some suitably constructive application of the ω-rule could yield a proof of a true Π_1-statement in the language of arithmetic not provable in, for example, Peano arithmetic, which could then be seen to be established *finitistically,* as the realization of Hilbert's Program calls for. Such a possibility is relevant to the present discussion insofar as a statement's being proved finitistically renders it arithmetical (as noted in Section I). Hence if, for example, the consistency statement for Peano arithmetic can be arrived at in these terms, it could not be correct to claim that Peano arithmetic is complete with respect to arithmetical truth.

Following Ignjatovic (1988), we may term what is provable by one application of the finitely expressed ω-rule extending finitistic arithmetic as "almost finitistically provable." Following Ignjatovic in following Tait (1981), we may identify finitistic provability with provability within primitive recursive arithmetic (PRA).[22] Accordingly, a Π_1-statement in the language of arithmetic is almost finitistically provable if it is provable in PRA extended by the rule

$$\frac{\mathrm{PRA} \;\vdash\; \forall z \exists y \mathrm{Prf}_{\mathrm{PRA}}(y, \mathrm{sub}(\ulcorner A(x) \urcorner, \mathrm{num}(z)))}{\forall x A(x).}$$

We must consider first the question whether this finitistically restricted ω-rule can be arithmetically justified (I shall argue that it can be) and then whether it properly extends Peano arithmetic (we shall see that it does not). The first step is to extend slightly the result from Feferman (1962:276, Theorem 2.19(i)) which we made use of in the previous section. Feferman states this result for Peano arithmetic and extensions of Peano arithmetic. The argument he gives goes through equally well, however, for weaker systems, in particular for primitive recursive arithmetic. Hence adding to PRA the effective ω-rule as formulated above yields a system which is equivalent to adding the uniform reflection principle (in either of its formulations) to PRA. PRA plus the uniform reflection principle over PRA is equivalent to extending PRA by the schema of transfinite induction of order type ω^ω. This result is the analog for PRA of the result from Kreisel and Levy (1968) cited in the previous section for

PA, that extension by uniform reflection and extension by ε_0 transfinite induction are equivalent. But transfinite induction of order type   is provable in Peano arithmetic. I have argued in Section I that everything provable in Peano arithmetic is arithmetical, so this last result tells us that almost finitistic provability (for statements of any logical complexity) is arithmetically justified. But at the same time, it also tells us that anything provable by these means lies *within* Peano arithmetic.[23]

It may seem that there is a form of the effectively applied   -rule lying midway between the case considered in the previous section and the case we have just now considered, namely extension of PA by the rule

$$\frac{PRA \vdash \forall z \exists y \, Prf_{PA}(y, sub(\ulcorner A(x) \urcorner, num(z)))}{\forall x A(x).}$$

This rule is, on the face of it, stronger than the rule we have considered in this section since it allows the substitution instances of the given formula to be proved in PA, rather than the more stringent condition that they must be proved in PRA. At the same time, it appears to be weaker than the rule of the previous section in that it has a seemingly stronger premise, namely that the condition that all substitution instances of the given formula be provable must itself be proved in PRA, rather than just in PA. So there is the *prima facie* possibility that it could still be arithmetically justifiable and yet prove something beyond Peano arithmetic. It does in fact prove something beyond Peano arithmetic, namely, for example, the Gödel sentence, as is easily verified by checking the proof using the corresponding form of   -rule given in the previous section (the properties of arithmetized syntax, including provable $_1$-completeness, required in the proof all hold for PRA as well as PA). However, this form of the rule is no more arithmetically justifiable than the form of the rule considered in the previous section, since in fact they are equivalent. Suppose that

$$PA \vdash \forall z \exists y \, Prf_{PA}(y, sub(\ulcorner A(x) \urcorner, num(z))).$$

Since $Prf_{PA}(x, y)$ is a primitive recursive predicate, there is a total recursive function f such that . But Nelson (1971) has extended Shoenfield's completeness result for the recursive   -rule to the primitive recursive   -rule. Modifying this proof, there exists a primitive recursive function h such that . The functions provably total in PRA ($_1$-PA) are precisely the primitive recursive functions. Hence

$$PRA \vdash \forall z \exists y \, Prf_{PA}(y, sub(\ulcorner A(x) \urcorner, num(z))).$$

It may appear that this last form of the rule can be strengthened to the previously considered and arithmetically justified:

$$\frac{PRA \vdash \forall z \exists y \, Prf_{PRA}(y, sub(\ulcorner A(x) \urcorner, num(z)))}{\forall x A(x).}$$

in the case when *"A(x)"* is a primitive recursive predicate. The idea would be that since PRA is complete with respect to true primitive recursive formulas, if then for *"A(n)"* a primitive recursive formula. Hence if num(z))), then But from the truth of this conditional, it does not follow that if PRA $\vdash$ then PRA This inference would hold if PRA $\vdash$ But then by univer-sal instantiation, and so, by contraposition, Since then But So then which is impossible by the Gödel theorem. Finally, we may note that the particular derivation in PA, and (as we have noted) also in PRA, given in Section V, of cannot be simply modified to yield a proof' of since we would then start with the assumption

(1)

and would need to modify the second step of the argument to

(2)

which would then yield

(3) Con(PRA).

But this is not adequate for the next step of the argument, which proceeds via

(4) Con(PA)

For the argument to carry on at this point we would need Con(PRA)   Con(PA), which we have noted cannot be proved in PA.

These last stages of the discussion may be taken to illuminate the sharpness of the boundary between arithmetical and non-arithmetical within the domain of statements in the language of arithmetic. The finitely expressed   -rule as an extension of Peano arithmetic, restricted to a primitive recursive formula, proves some but not all true $_1$-statements not provable in PA, but we also saw that it is not arithmetically justifiable. The finitely expressed   -rule as an extension of primitive recursive arithmetic is arithmetically justifiable but also fails to prove anything not already provable in Peano arithmetic. We then considered hybrids of these two forms of the   -rule, potentially intermediate between them, and found that in each case the result fell decisively on one side or the other of the divide between these two forms.

VII.
CONCLUDING REMARKS

The aim of this paper has been to test the thesis that Peano arithmetic is complete with respect to an epistemic notion of arithmetical truth by considering the nature of the ω-rule in proving truths beyond Peano arithmetic. The question has been whether the ω-rule, either in its basic form or in some restricted more constructive version, can be given arithmetical justification, in the sense being considered here, and in a case where it can be, whether it provides derivations of truths in the language of arithmetic not provable in Peano arithmetic.

Sections I and II were introductory. In Section III we considered an argument based on the close connection between the ω-rule and the clause for the universal quantifier in an inductive definition of truth over the structure of the natural numbers which appeared to provide arithmetical justification to the ω-rule. The key element of this argument was the global reflection principle for Peano arithmetic, that everything provable in Peano arithmetic is true in the structure of the natural numbers. I argued that its justification from our basic conception of the structure of the natural numbers requires a second-order process of reflection which renders it non-arithmetical in the sense at issue. We saw also that uniform reflection could be used in this argument but that the extension from what was clearly arithmetically justified to the required principle was precisely an instance of the ω-rule. We also noted that while the principle of induction is justifiable from the ω-rule, the justification itself requires use of induction, so that the ω-rule was not thereby shown to be characteristic of our conception of the structure of the natural numbers.

In Section IV we turned to justification of the ω-rule from ω-consistency. A theory which is true in the structure of the natural numbers is ω-consistent, but the considerations of the preceding section showed that the ω-consistency of Peano arithmetic could not be established arithmetically in this way. We noted then that, while being true in the natural numbers is a sufficient condition for ω-consistency, it is not necessary, and we turned to consider other ways of establishing this property, and then also considered establishing just the 1-consistency of PA, as a means of validating the ω-rule as restricted to primitive recursive formulas. We noted that this could be done by use of transfinite induction of order type ϵ_0, and I then offered arguments that transfinite induction of order type ϵ_0 cannot be justified arithmetically. While no weaker form of transfinite induction can establish the 1-consistency of PA, ϵ_0 transfinite induction is stronger than the 1-consistency of PA in that it proves reflection principles stronger than the 1-consistency of PA. The 1-consistency of PA is equivalent to Σ_1-local reflection, which in turn is equivalent to the Paris-Harrington sentence. While it is not possible to say definitively that the latter statements cannot be proved arithmetically, the given proofs of the Paris-Harrington sentence arguably are not, and the strong soundness property of Σ_1-reflection suggests that considerations akin to those in Section III will apply here also.

In Section V we considered the extension of Peano arithmetic by the finitely applied ω-rule. We noted that extension of Peano arithmetic by this form of the ω-rule is equivalent to extension of Peano arithmetic by ϵ_0 transfinite induction, seen by earlier arguments not to be arithmetically justifiable. There then remained the possibility that a restricted form of the finitely applied ω-rule as an extension of Peano arithmetic might be arithmetically justified, and we considered the restriction to primitive recursive formulas. We saw by an example that this case is weaker than the ordinary ω-rule restricted to primitive recursive formulas, which we had considered in the previous section, but that it is still strong enough to generate a proper extension to Peano arithmetic by deriving the Gödel sentence. We then saw that justification of this form of the rule is tantamount to establishing the consistency of Peano arithmetic. Hence its use to establish the Gödel sentence, which is equivalent to Con(PA), cannot represent a gain in arithmetically justified truth.

The final step was to consider, in Section VI, the finitely applied ω-rule as an extension of finitistic arithmetic, which we took as primitive recursive arithmetic. We noted that extension by this form of the rule is tantamount to transfinite induction of order type ω, which is arithmetically provable, but thereby everything provable with this form of the rule is already provable in Peano arithmetic.

I conclude that the investigation we have followed here shows the stability of this notion of arithmetical truth and its coincidence with Peano arithmetic. The ω-rule, which appears initially as a powerful tool for extending beyond provability in Peano arithmetic, turns out to require for its justification, so far as the rule does extend Peano arithmetic, acceptance of modes of argument which themselves lie beyond arithmetical truth. As we weaken the rule, we find that the justification required becomes comparable with what can be achieved with the rule, and when finally we have weakened it sufficiently to achieve arithmetical justification, what it then proves falls within Peano arithmetic.

These considerations encourage the view that we have identified a natural type within the space of mathematical knowledge. The enterprise here might be compared with the identification of finitistic mathematics with primitive recursive arithmetic (Tait 1981), and of predicative with hyperarithmetic (Kreisel 1960, 1962; Feferman 1964).[24]

I conclude with the observation that this paper answers a question raised by Tarski (1936). Following discussion of the nature of the ω-rule (infinite induction), Tarski considers (p. 412) that this rule

does not deviate essentially from the [ordinary] rules of inference, either in the conditions of its applicability or in the nature of the concepts involved in its formulation or, finally, in its intuitive infallibility (although it is considerably more complicated),

and he then remarks that

It would be interesting to investigate whether there are any objective reasons for assigning a special position to the rules ordinarily used.

The present paper provides an affirmative answer to Tarski's question.

NOTES

1　I will talk about "*the* language of arithmetic" elliptically for "the *chosen* (first-order) language of arithmetic" (which might, for given purposes, include, besides the constants 0, S, +, •, symbols for other or even all primitive recursive functions, and/or various predicate symbols).

2　This expression of our grasp of the structure of the natural numbers was set out by Dedekind (1888) (see also his letter to Keferstein (1890)). Frege's (1884, 1879) treatment of the natural numbers is mathematically equivalent to Dedekind's, though somewhat differently motivated and developed.

3　This formulation by Gödel must have been a key source for me of this notion, as I used it in 1987, even though at the time of writing that paper I seem not to have had it explicitly in mind.

4　Hilbert's rule was taken up by Jacques Herbrand (1931:624) (Group D of his axioms of arithmetic). The constructive restrictions implicit in Hilbert's formulations are spelled out by Herbrand. (Herbrand's paper is dated "Göttingen, 14 July 1931." Hilbert presented his paper (1931b) to the Gesellschaft der Wissenschaften zu Göttingen on 17 July 1931. Herbrand's fatal mountaineering accident occurred on 27 July 1931.)

5　He there talks about sentences of his Language I being analytic or contradictory, but that dichotomy for Carnap (1934) corresponds exactly to true or false in the structure of the natural numbers.

6　See Schütte (1951: he there cites Hilbert (1931a); and 1960:168); for a more recent account see Schwichtenberg (1977).

7　At one point, Tarski (1933:279, fn 2) refers to the -rule as "the rule of transfinite induction." I assume this is a verbal slip, where he meant to have written his usual designation of the rule as "infinite induction," and should not be taken as echoing this usage by Hilbert. (Tarski cites Hilbert (1931a), which does not contain this usage, and makes no reference to (1931b), which does, when he remarks, in footnote 1 on p. 295, that "the rule of inference mentioned has recently been discussed by D.Hilbert".)

8　The similarity which is implicit in Carnap's discussion in the passage in question is that these rules are all cited as not being d-rules, "d" for derivation, to which most systems are restricted, but are c-rules, that is, rules of consequence—see Carnap (1937:39) for discussion of the distinction between derivation and consequence.

9　There is similarly ambiguous juxtaposition on p. 197, where Carnap notes that, with his rule is a consequence of the collection of sentences *A(t),* and then goes on to observe that "in the other languages which we have mentioned, on the contrary, the same thing is not true for the universal operators… (unless Hilbert's new rule is laid down)," again refraining from drawing an explicit connection between his rule DC2 and Hilbert's rule.

10　Compare the opening lines of Frege (1918): "The word 'true' indicates the aim of logic as does 'beautiful' that of aesthetics or 'good' that of ethics. All sciences have truth as their goal; but logic is also concerned with it in a quite different way from this. It has

much the same relation to truth as physics has to weight or heat. To discover truths is the task of all sciences; it falls to logic to discern the laws of truth."

11 Richard Shore noted in discussion with me the commitment to general properties of truth which follows from my characterization of arithmetical truth as closed under logical deduction (though the context of his point was somewhat different than it is here). Sundholm (1978:1) considers the basic ω-rule to be "nothing but the $\forall$-clause in the (substitutional) truth-definition for [sentences in the language of arithmetic]." However, we shall see now in further discussion of the preceding derivation that the ω-rule and the clause for the universal quantifier in an inductive definition of truth, while closely related, are not identical.

12 It has indeed been claimed that the ω-rule is more basic, or elementary, than the principle of induction. Lopez-Escobar (1976) opens his paper with the remark that "from an intuitive point of view the ω-rule [he gives it as we had it at the beginning of Section II] is a much simpler rule to justify than its finitary cousin, the rule of induction: From $A(0)$ and to conclude ."

13 Tarski (1933b: 288–9) also shows how to construct examples of theories which are both consistent and ω-inconsistent (which he claims in footnote 2, p. 279, he had already presented in a lecture in 1927, while eschewing any claim that he "already knew then the results later obtained by Gödel or had even foreseen them").

14 Compare this entailment with the remark by Tarski (1936: p. 258, fn 1) in which he equates the ω-consistency of a class of sentences in a language for arithmetic with the condition that this class remains consistent after a single application of the ω-rule (infinite induction).

15 A primitive recursive formula is said to be both Σ_0 and Π_0. A formula A is said to be Σ_{n+1} if there is a Π_n-formula B such that A is equivalent to , and A is Π_{n+1} if there is a Σ_n-formula B such that A is equivalent to (cf. Smorynski 1977:843).

16 If , then since T has an ω-model, is true in the structure of the natural numbers, which is to say that, for some natural number is true in the natural numbers, that is, $A(\underline{k})$ is not true, which means that .

17 As shown by the following example: Let T be the theory PA+ arithmetization of "this sentence, when added to PA, produces a system which is ω-inconsistent." The following is a heuristic argument to show that T is ω-consistent and has no ω-model. If the sentence added to PA did produce an ω-inconsistency, it would be true (in the structure of the natural numbers). But in that case it would not have produced an ω-inconsistency when added to PA, by the argument in note 14. So it must be that it does not produce a ω-inconsistency when added to PA, which is to say that the theory in question must be ω-consistent. But then the added sentence is false, so the whole theory cannot be true in the natural numbers, that is, has no ω-model (Kreisel 1955).

18 A reflection principle for a given system T is an assertion of the soundness of that system which in global terms would be the statement

For every sentence A in the language of T, if A is provable in T, then A is true.

(Compare supposition 1 of Section III.) For any theory T to which Tarski's theorem on the undefinability of truth is applicable, such a global reflection principle cannot be expressed in the language of T. We can express soundness of T locally in the language of T by a *schema* using the proof relation $\mathrm{Prf}_T(x,y)$ from arithmetized syntax, and where $\ulcorner A \urcorner$ signifies the Gödel number of the formula "A," namely:

$$\exists y \mathrm{Prf}_T(y, \ulcorner A \urcorner) \rightarrow A,$$

for each sentence (closed formula) in the language of T.

While global reflection for T cannot be expressed in the language of T, it is possible to parametrize infinite classes of sentences of the form $\{A(0), A(S0), ..., A(n),...\}$ by use of a universally quantified variable, resulting in this version of uniform reflection, where sub(u, v) is the primitive recursive function of arithmetized syntax which yields the numerical code of the formula which results from substituting the term whose code is v in place of the (only, or specified) free variable in the formula whose code is $u,$ and num(z) gives the code of the formal numeral for the natural number z:

$$\forall z [\exists y \mathrm{Prf}_T(y, \mathrm{sub}(\ulcorner A(x) \urcorner, \mathrm{num}(z))) \rightarrow A(z)]$$

(compare supposition 9 of Section III). For Σ_1-statements local reflection and uniform reflection over Peano arithmetic are equivalent (see Smorynski (1977:846, Theorem 4.1.4)), though in the general case uniform reflection is strictly stronger than local reflection (Feferman 1962:276, Theorem 2.19(ii)).

Reflection for a formal system is in general a stronger condition of soundness than consistency (heuristically, someone can be untruthful but still be consistent). Local reflection for Σ_1-formulas does not follow from consistency (Feferman 1962:274–5, Lemma 2.17(ii)) since, in particular,

$$\mathrm{PA} + \mathrm{Con}(\mathrm{PA}) \nvdash \mathrm{Prov}_{\mathrm{PA}}(\ulcorner \neg \mathrm{Con}(\mathrm{PA}) \urcorner) \rightarrow \neg \mathrm{Con}(\mathrm{PA}),$$

by the Gödel Second Incompleteness Theorem. However, local reflection for Σ_1-formulas is implied by consistency (Hilbert (1928:474) sketches the essential idea for this implication, which lies at the heart of Hilbert's Program). Global reflection is stronger than consistency already for Σ_1-statements, since for G the Gödel sentence, is consistent but does not satisfy global reflection since ¬G is false.

19 I am grateful to Aleksandar Ignjatovic and Robin Gandy for kindly explaining this result to me.

20 I am grateful to Richard Kaye for this example.

21 This case of the ω-rule is studied by Aleksandar Ignjatovic (1988). My discussion here is dependent on his illuminating ideas and results.

22 PRA is a formalization of the quantifier-free arithmetic established by Skolem (1923), whose language must contain a primitive function symbol for each primitive recursive function (see for example Girard (1987:66–7) for such a formalization). For present

purposes PRA is comparable to a system of Peano arithmetic in which the induction schema is restricted to ₁-formulas, which I will here term ₁-PA, since the provably total functions of ₁,-PA are precisely the primitive recursive functions (see Wainer 1990), so that *yA(x, y)* if and only if there is a primitive recursive function *f* such that PRA . In the following discussion I shall continue to speak of PRA, as it is conceptually, in terms of finitistic constructivity, the more fundamental system (also it is neater to print), but I will mean ₁-PA, since that system allows expressions with quantifiers.

23   While almost finitistic provability via this effective use of the -rule does not constitute an extension of Peano arithmetic, it does give a proper extension of PRA. In particular, Con(PRA), by Gödel's Second Incompleteness Theorem not provable in PRA, is almost finitistically provable (see Ignjatovic (1988:13, Corollary 1)); this result corresponds to the provability of the Gödel sentence for PA (and similarly for Con(PA)) by use of the arithmetized -rule for PA, which we looked at in the latter part of Section V).

24   There is a difference in the historical situation of the present project from the two just cited, namely that in those cases the conceptual notions pre-dated in philosophical discussion the systems offered as their precise realizations. The notion of arithmetical, which here and in Isaacson (1987) I have argued is to be identified with Peano arithmetic, does not seem previously to have been singled out for consideration.

REFERENCES

Quotation and page references are to the last indicated reprinting or translation; sometimes page reference to the original printing is also given.

Bernays, P. (1935) "Sur le platonisme dans les mathematiques," *L Enseignement Mathematique* 34:52–69; English translation by C. Parsons, "On platonism in mathematics," in P.Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings,* 2nd edn, Cambridge: Cambridge University Press, 1983, 258–71.

Cantor, G. (1883) "Grundlagen einer allgemeinen Mannigfaltigkeitslehre," No. 5 of the series *Über unendliche lineare Punktmannigfaltigkeiten, Mathematische Annalen* 21:545–86, reprinted in E.Zermelo (ed.) *Georg Cantor Gesammelte Abhandlungen,* Berlin: Springer, 1932, 165–209.

Carnap, R. (1934) *Logische Syntax der Sprache,* Vienna: Julius Springer, 274pp.

——(1935) "Ein Gültigkeitskriterium für die Sätze der klassischen Mathematik," *Monatshefte für Mathematik und Physik* 42:163–90.

——(1937) *The Logical Syntax of Language,* London: Routledge & Kegan Paul, 352pp., English translation by A.Smeaton, with additions, of *Logische Syntax der Sprache,* Vienna: Julius Springer, 1934.

——(1963) "Intellectual autobiography," in P.A.Schilpp (ed.) *The Philosophy of Rudolf Carnap,* The Library of Living Philosophers, La Salle, IL: Open Court, 1–84.

Dedekind, R. (1888) *Was sind und was sollen die Zahlen?,* Braunschweig: Vieweg; trans. W.W.Beman, "The nature and meaning of numbers," in R.Dedekind, *Essays on the Theory of Numbers,* Chicago, IL: Open Court, 1901; reprinted New York: Dover, 1963, 31–115.

——(1890) Letter to Keferstein, 27 February 1890; English translation by H.Wang and S.Bauer-Mengelberg, in J.van Heijenoort (ed.) *From Frege to Gödel: A Sourcebook in Mathematical Logic 1879–1931,* Cambridge, MA: Harvard University Press, 1967, 98–103.

Feferman, S. (1962) "Transfinite recursive progressions of axiomatic theories," *Journal of Symbolic Logic* 27:259–316.

——(1964) "Systems of predicative analysis," *Journal of Symbolic Logic* 29:1–30.

——(1986) Introductory note to "Review of Hilbert, 'Die Grundlegung der elementaren Zahlenlehre,'" *Kurt Gödel Collected Works,* vol. 1, Oxford: Oxford University Press, 208–13.

Feferman, S. and Spector, C. (1962) "Incompleteness along paths in progressions of theories," *Journal of Symbolic Logic* 27:383–90.

Frege, G. (1884) *Die Grundlagen der Arithmetik; Eine logisch mathematische Untersuchung über den Begriff der Zahl,* Breslau: Wilhelm Koebner, 119pp.; English translation by J.L.Austin, *The Foundations of Arithmetic; a logico-mathematical enquiry into the concept of number,* Oxford: Basil Blackwell, 1950, 119pp.

——(1918) "Der Gedanke: eine logische Untersuchung," *Beiträge zur Philosophic des deutschen Idealismus* 1:58–77; English translation by A.M.Quinton and M.Quinton, "The thought: a logical inquiry," *Mind* 65 (1956): 289–311.

Girard, J.-Y. (1987) *Proof Theory and Logical Complexity,* vol. 1, Naples: Bibliopolis, 503pp.

Gödel, K. (1931a) "Über formal unentscheidbare Sätze der *Principia mathematica* und verwandter Systeme I," *Monatshefte für Mathematik und Physik* 38:173–98; reprinted with facing translation by J.van Heijenoort, "On formally undecidable propositions of *Principia Mathematica* and related systems I," in S.Feferman *et al.* (eds) *Kurt Gödel Collected Works,* vol. 1, Oxford University Press, 1986, 145–95.

——(1931b) Review of Hilbert "Die Grundlegung der elementaren Zahlenlehre," *Zentralblatt für Mathematik unde ihre Grenzgebiete 1:* 260; reprinted with facing English translation in S.Feferman *et al.* (eds) *Kurt Gödel Collected Works,* vol. 1, Oxford: Oxford University Press, 1986, 212–15.

——(1958) "Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes," *Dialectica* 12:280–7; reprinted with English translation by S.Bauer-Mangelberg and J.van Heijenoort in S. Feferman *et al.* (eds) *Kurt Gödel Collected Works,* vol. 2, Oxford: Oxford University Press, 1990, 240–1.

——(1972) "On an extension of finitary mathematics which has not yet been used," revisions and expansions by Gödel of an English translation by L.F.Boron of "Uber eine bisher noch nicht benützte Erweiterung des finiten Standpunktes,' in S.Feferman *et al.* (eds) *Kurt Gödel Collected Works,* vol. 2, Oxford: Oxford University Press, 1990, 271–80.

Grzegorczyk, A., Mostowski, A. and Ryll-Nardzewski, C. (1958) "The classical and the   -complete arithmetic," *Journal of Symbolic Logic* 23:188–206.

Herbrand, J. (1931) "Sur le non-contradiction de l'arithmetique," *Journal für die reine und angewandte Mathematik* 166:1–8; English translation and introduction by J.van Heijenoort, "On the consistency of arithmetic," in J.van Heijenoort (ed.) *From Frege to Gödel: A Sourcebook in Mathematical Logic 1879–1931,* Cambridge, MA: Harvard University Press, 1967, 618–28.

Hilbert, D. (1925) "Über das Unendliche," *Mathematische Annalen* 95: 161–90; trans. S.Bauer-Mengelberg, "On the infinite," in J. van Heijenoort (ed.) *From Frege to Gödel: A Sourcebook in Mathematical Logic 1879–1931,* Cambridge, MA: Harvard University Press, 1967, 367–92.

——(1928) "Die Grundlagen der Mathematik," *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität* 6:1–21; English translation by S.Bauer-Mengelberg and D.Føllesdal, "The foundations of mathematics," in J.van Heijenoort (ed.) *From Frege to Gödel: a source book in mathematical logic 1879–1931,* Cambridge, MA: Harvard University Press, 1967, 464–79.

——(1931a) "Die Grundlegung der elementaren Zahlenlehre," *Mathematische Annalen* 104:485–94.

——(1931b) "Beweis des Tertium non datur," *Nachrichten der Gesellschaft der Wissenschaften zu Göttingen:* 120–5.

Ignjatovic, A. (1988) "Hilbert's program and the   -rule," Unpublished, Group in Logic and Methodology of Science, University of California at Berkeley, 31pp.

Isaacson, D. (1987) "Arithmetical truth and hidden higher-order concepts," in *Logic Colloquium '85,* edited by the Paris Logic Group, Amsterdam: North-Holland, 147–69. [*Errata:* p. 152, line 10 up, should read $\forall X(\forall y(\forall x(x \leqslant y \rightarrow x \in X) \rightarrow y \in X) \rightarrow \forall x(x \in X))$; p.162, top line, should read "the cardinality of the set should be greater than the least number in the set."]

Kaye, R. (1989) "Model-theoretic properties characterizing Peano Arithmetic," Unpublished, Jesus College, Oxford, 27pp.

Kreisel, G. (1955) Review of Leon Henkin, "A generalization of the concept of   -consistency" (*Journal of Symbolic Logic* 19 (1954): 183–96), *Mathematical Reviews* 16:103.

——(1960) "La predicativité," *Bulletin de la Société mathématique de France:* 371–91.

——(1962) "The axiom of choice and the class of hyperarithmetic functions," *Koninklijke Nederlands Akademie van Wetenschappen, Proceedings, Series A* 65 (also *Indagationes Mathematicae* 24): 307–19.

Kreisel, G. and Levy, A. (1968) "Reflection principles and their use for establishing the complexity of axiomatic systems," *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* 14:97–142.

Lopez-Escobar, E.G.K. (1976) "On an extremely restricted   -rule," *Fundamenta Mathematicae* 90:159–72.

Mostowski, A. (1947) "On definable sets of positive integers," *Fundamenta Mathematicae* 34:81–112.

——(1966) *Thirty Years of Foundational Studies; Lectures on the development of mathematical logic and the study of the foundations of mathematics in 1930–1964, Acta Philosophica Fennica* 17 and Oxford: Basil Blackwell, 180pp.

Nelson, G.C. (1971) "A further restricted   -rule," *Colloquium Mathematicum* 23:1–3.

Paris, J. and Harrington, L. (1977) "A mathematical incompleteness in Peano Arithmetic," in J.Barwise (ed.) *Handbook of Mathematical Logic,* Amsterdam: North-Holland, 1133–42.

Rosser, B. (1937) "Gödel theorems for non-constructive logics," *Journal of Symbolic Logic 2:*129–37.

Schütte, K. (1951) "Beweistheoretische Erfassung der unendlichen Induktion in der Zahlentheorie," *Mathematische Annalen* 122:369–89.

——(1960) *Beweistheorie,* Berlin: Springer Verlag, 355pp.

——(1977) *Proof Theory,* Berlin: Springer Verlag, 299pp.

Schwichtenberg, H. (1977) "Proof theory: some applications of cutelimination," in J.Barwise (ed.) *Handbook of Mathematical Logic,* Amsterdam: North-Holland, 867–95.

Shoenfield, J.R. (1959) "On a restricted   -rule," *Bulletin de l'Academie Polonaise des Sciences 7:*405–7.

Skolem, T. (1923) "Begründung der elementaren Arithmetik durch die rekurrierende Denkweise ohne Anwendung scheinbarer veränderlichen mit unendlichem Ausdehnungsbereich," *Videnskapsselskapets skrifter, I. Matematikshaturvidenskabelig klasse* 6: 38pp.; English translation by S.Bauer-Mengelberg, "The Foundations of elementary arithmetic established by means of the recursive mode of thought, without the use of apparent variables ranging over infinite domains," in J.van Heijenoort (ed.) *From Frege to Gödel: A Sourcebook in Mathematical Logic 1879– 1931,* Cambridge, MA: Harvard University Press, 1967, 303–33.

Smorynski, C. (1977) "The incompleteness theorems," in J.Barwise (ed.) *Handbook of Mathematical Logic,* Amsterdam: North-Holland, 821–65.

Sundholm, G. (1978) "The Omega-Rule: A Survey," B.Phil, thesis, Oxford University, 102pp.

——(1983) "Proof Theory: a survey of the omega-rule," D.Phil, thesis, Oxford University, 173pp.

Tait, W.W. (1981) "Finitism," *Journal of Philosophy* 78:524–46. Tarski, A. (1933a) "Pojecie prawdy w jezykach nauk dedukcyjnych," *Prace Towarzystwa Naukowego Warszawskiego, wydzial III* no. 34; German translation by L.Blaustein, with additional material, "Der Wahrheitsbegriff in den formalisierten Sprachen," *Studio Philosophica* 1 (1936): 261–405; English translation by J.H.Woodger of the German version, "The concept of truth in formalized languages," *Logic, Semantics, Metamathematics: papers from 1923 to 1938 by Alfred Tarski,* Oxford: Oxford University Press, 1956; 2nd edn, with corrections and emendations, ed. by John Corcoran, Indianapolis, IN: Hackett, 1983, 152–278.

——(1933b) "Einige Betrachtungen über die Begriffe   -Widerspruchsfreiheit und der   -Vollständigkeit," *Monatshefte für Mathematik und Physik* 40; English translation by J.H.Woodger as "Some observations on the concepts of   -consistency and   -completeness," *Logic, Semantics, Metamathematics: papers from 1923 to 1938 by Alfred Tarski,* Oxford: Oxford University Press, 1956; 2nd edn, with corrections and emendations, ed. by John Corcoran, Indianapolis, IN: Hackett, 1983, 279–95.

——(1936) "Über den Begriff der logischen Folgerung," *Actes du Congres International de Philosophic Scientifique 7 (Actualites Scientifique et Industrielles 394),* Paris, 1936, 1–11; English translation by J.H. Woodger, "On the concept of logical consequence," *Logic, Semantics, Metamathematics: papers from 1923 to 1938 by Alfred Tarski,* Oxford: Oxford University Press, 1956; 2nd edn, with corrections and emendations, ed. by John Corcoran, Indianapolis, IN: Hackett 1983, 409–20.

Wainer, S.S. (1990) "Computability—logical and recursive complexity," Preprint 31/90, Centre for Theoretical Computer Science, University of Leeds, 28pp.; to appear in S.L.Bauer (ed.) *Proceedings of the NATO Summer School on Logic, Algebra and Computation at Marktoberdorf, 1989,* Berlin: Springer-Verlag.

Wilkie, A.J. (1981) "On discretely ordered rings in which every definable ideal is principal," in C.Berline, K.McAloon, and J.-P.Ressayre (eds) *Model Theory and Arithmetic (Proceedings, Paris, 1979/80),* Springer Lecture Notes in Mathematics 890, Berlin: Springer, 297–303.

——(1987) "On schemes axiomatizing arithmetic," *Proceedings of the International Congress of Mathematicians, Berkeley, 1986,* pp. 331–7.

6
THE IMPREDICATIVITY OF INDUCTION
Charles Parsons

SUMMARY

It has been frequently objected to the reduction of mathematical induction to a definition first proposed by Frege that it is impredicative in that its development requires second-order logic or set theory with impredicative comprehension assumptions. The present paper develops an observation of Michael Dummett that a similar impredicativity remains if induction is treated not as a definition but as integral to an informal explanation of the predicate "natural number" *(N),* even though full second-order logic may not be used, for example if we think of the introduction of *N* as an inductive definition.

The argument for this claim applies to the case of natural numbers a well-known one for the impredicativity of "generalized inductive definitions" such as Kleene's definition of the class *O* of recursive ordinal notations. Such generalized inductive definitions are impredicative according to the persuasive analysis of predicativity given by Solomon Feferman and Kurt Schütte.

The observation that the notion of natural number is already impredicative is not an objection to this analysis, which was intended to capture the notion of predicativity *given* the natural numbers. But it seriously weakens the case for the claim, deriving from Poincaré', that impredicativity is a sign of a vicious circle and altogether to be avoided.

Poincaré and others who followed him assumed a conception of sets or classes as extensions of predicates antecedently understood. It is argued that the use of generalized inductive definitions by constructivists such as Lorenzen is consistent with this conception; in this sense it is predicative, although not in the Feferman-Schütte sense.

We consider very briefly the objection that the claim that the notion of natural number is impredicative relies on a questionable notion of conceptual truth.

According to Gottlob Frege and many logicists since his time, the principle of mathematical induction is a consequence of the definition of natural number. Let *"Na"* mean "*a* is a natural number" and let *A(x)* be an open sentence. Then we can take the principle as a rule of inference that enables us to go from the premises

$$A(0)$$
$$Na \rightarrow [A(a) \rightarrow A(Sa)]$$
$$Nt$$

to the conclusion *A(t),* for any term *t.* In effect, Frege defined *"Na"* as

$$\forall F\{[F0 \wedge \forall x(Fx \rightarrow F(Sx))] \rightarrow Fa\}$$

and then, given his logic (a form of second-order logic), the principle of induction is indeed a derived rule.

A frequently voiced objection to Frege's view of induction is that the application of his definition requires the second-order logic to be impredicative. It must allow the instantiation of predicates containing second-order quantifiers for *F* in the definition of *"N."* We can see this if we reflect that, in order to carry out even elementary proofs in number theory, we need to apply induction in cases where the predicate *A(x)* contains the predicate *"N,"* which, *ex hypothesi,* has been defined by second-order quantification.

I shall use the term "second-order entity" as a neutral term for whatever our second-order variables range over. Different interpretations will make them Fregean concepts, prepositional functions, sets, classes, or attributes. The point is that a logic in which Frege's definition yields the instances of induction needed even for elementary arithmetic must allow instantiation for a universal quantifier of predicates containing quantification over all secondorder entities; thus these entities are taken to include some defined in terms of quantification over all of them, a totality to which they themselves belong.[1]

Different attitudes have been taken concerning the force of this objection, depending on one's view of impredicativity in general. The most widely held view is perhaps that it is an objection to taking full second-order logic to be truly *logic,* in the sense of consisting of rules that are basic to all reasoning about objects in general. Such second-order logic includes,

implicitly or explicitly, assumptions as to when a predicate expresses or stands for a second-order entity. In particular, it assumes that predicates containing quantifiers over such entities do so. Such assumptions are in effect comprehension axioms and have an existential character; they are not essentially different from existence axioms for sets or classes.

The thesis of the present note is that the impredicativity that arises from Frege's attempt to reduce induction to a definition is not a mere artifact of Frege's strategy of reduction. As Michael Dummett observed some years ago, the impredicativity—though not necessarily impredicative second-order logic—remains if we regard induction in a looser way, as part of the explanation of the term "natural number."[2] If one explains the notion of natural number in such a way that induction falls out of the explanation, then one will be left with a similar impredicativity. The same holds for other domains of objects obtained by iteration of operations yielding new objects, beginning with certain initial objects. It seems that the impredicativity will lose its significance only from points of view that leave it mysterious why mathematical induction is evident. A conclusion I wish to draw from this simple observation is that some impredicativity is inevitable in mathematical concept formation.[3] This contradicts the once-influential view deriving from Poincaré that it is a clear sign of a vicious circle. At the end of the paper I will comment on some recent views of impredicativity, particularly those of Lorenzen and Feferman.

<div align="center">I</div>

Induction appears to be implicit in our concept of natural number, but there are other ways of capturing this idea than Frege's explicit definition. One way that avoids some of the difficulties of Frege's is viewing the predicate "natural number" as introduced by an inductive definition. Some would find the term "definition" inappropriate here, and it is not essential. What we have is a system of rules which serves to explain a newly introduced predicate. Unlike an explicit definition, these rules do not lead to the eliminability in principle of the predicate from contexts in which it occurs, but other arguments can be given for their adequacy. We will assume that we understand the term "0" and an operation S leading from x to its successor Sx. Terms of the form *"SS... S0"* behave like singular terms, but I do not wish to take this too strictly, since I want my remarks to apply to a situation where variables for numbers are substitutional. Moreover, where such variables are understood objectually, nothing is to turn on any claim to the effect that what these terms denote are really natural numbers rather than elements of some other model of arithmetic. My remarks should apply to any instance of the structure of natural numbers, in particular intuitive models such as Hilbert's strings of strokes.[4] This is in keeping with the widely held view that the natural numbers are given only by their structure.

Our conception of the natural numbers is of what is obtained by beginning with 0 and iterating the successor operation. Frege tried to capture this by an explicit definition; the present approach undertakes to capture it by rules. Obviously it allows us to assert that 0 is a natural number and that Sx is a number if x is. Hence we have the following "introduction rules":

$$N0$$

$$\text{From } Nx \text{ infer } N(Sx).$$

Now these rules are understood as the canonical way of arriving at statements to the effect that something is a natural number; it is only by virtue of them that something is a natural number. Thus a common way of stating inductive definitions is by giving rules like these and then saying something like, "Nothing is a number except by these rules." That amounts to taking N to be *minimal* so that the introduction rules hold. But now suppose $A(a)$ is a predicate for which the introduction rules hold:

$$A(0)$$

$$A(x) \rightarrow A(Sx).$$

Then $A(a)$ must be true of any natural number. But that is just the induction principle. It can be given the form of an "elimination rule" for N, that is, a rule of inference that from the above two premises with the additional premise Nt allows inference of $A(t)$.[5]

There is a very natural picture that arises here. If x is a number, then by the introduction rules, if we begin with 0, then by a succession of steps from y to Sy we reach x. Then by a parallel succession of steps, we can show that $A(y)$ holds for each y figuring in the construction, and therefore that $A(x)$ holds. In fact, for each x we can construct a formal proof of $A(x)$ by beginning with $A(0)$ and building up by *modus ponens,* using $A(x)$ $A(Sx)$. As a *proof* of induction, this is circular: the "construction" of x by a *succession* of steps is itself inductively defined, and it is by a corresponding induction that it is established that A holds at each point in the construction. Nonetheless, it is still useful for meta mathematical arguments concerning induction in formalized theories, and it is no worse than arguments for the validity of elementary logical rules.

Now evidently the state of affairs that made Frege's definition require impredicative logic also obtains in the present setting: in order to apply the induction rule to prove significant generalizations about numbers, we will have to use predicates containing the predicate *"N."* As originally offered, as a principle cashing in our intention that the numbers should be what is obtained by the introduction rules and those alone, the principle refers to arbitrary predicates, without any assumptions having

‖‖‖‖‖‖‖‖‖.

been made about what counts as a predicate. Like the principles of predicate logic itself, we have a purely formal generalization about predicates, which is not a generalization over a given *domain* of entities and could not be, since it is not determined what predicates will or can be constructed and understood. If our explanation of the natural numbers is successful, then we do understand *"N"* as a predicate and indeed as describing a possible domain for quantification; that is, we are able to understand quantifiers restricted to *N*. But this means that the impredicativity of Frege's definition survives in the present setting, since we have explained *"N"* in part by a formal generalization about predicates, which has to admit as instances predicates containing *"N"* itself.

Our situation can be described as follows. The readily available alternatives to something like the inductive-definition model of the concept of natural number (with or without the Fregean reduction) would be to give it an explanation that is blatantly circular, such as, the natural numbers are what is obtained by beginning with 0 and iterating the successor operation an *arbitrary finite number* of times, or to take the concept of natural number as given and the principle of induction as evident without any explication connecting it with the concept of natural number. Either alternative seems to me a counsel of philosophical despair that leaves us with no motivation for the principle of induction. Below, however, we will consider some alternatives from set theory. The explanation we have includes a generalization about predicates that includes the predicate *"N"* in its scope.

In interpreting both Frege's definition and our characterization by rules of the predicate *"N"*, one can ask what is the range of the *first*-order variables. In some applications of inductive definitions, all the definition aspires to do is to define a predicate of objects of a *previously* given domain and so we can take first-order quantifiers to be understood in some independent way. Frege himself seems to have assimilated his own case to this one in assuming that first-order quantification is quantification over *all* objects, in an absolute sense independent of the particular context of inquiry. Such a view is not very persuasive now. Should we still interpret our explanation of the notion of natural number as picking out the natural numbers from a previously given domain? One might reply that our conception of the natural numbers is that of a *structure* and therefore does not give individual identities to the objects playing the role of 0, 1, 2, etc.; therefore there should be no unique answer to the question from what domain the natural numbers are picked out or even whether there is one. The generality with which we have proceeded is in keeping with this structuralist view and therefore cannot exclude the case where there is a previously given domain. Some instances of the structure of natural numbers are substructures of others, and we might describe such substructures by inductive definitions.

However, to assume that this is always the case is to assume that some infinite structure is given to us independently of our knowledge of the kind of structure the natural numbers instantiate. This may be a domain of mathematical objects such as sets. Or one may, as in nominalist views, claim that one can find in the physical world a realization of the structure of natural numbers, which is therefore to be described by quantifiers ranging over some domain of physical objects. Either of these alternatives has its difficulties, which it would take us too far to go into.[6] My main point concerning the impredicativity of induction is independent of whether the natural numbers (or a representative of their structure that is of some basic significance) are a class of previously given objects.

If there is no such previously given infinite structure, then it is as if we had arrived at the concept of natural number by pulling ourselves up by our conceptual bootstraps, so as to understand the notion of some such structure and convince ourselves of its possibility without having in advance the conception of a domain of objects from which the objects of the structure are picked out.

In my view this is possible for an intuitive model such as the Hilbertian one. Let us suppose for the moment that "0" denotes the symbol "‖" (as a type, understood as a spatial configuration), and that, from a given *x, Sx* is the result of adding another "‖" on the right. Then the objects of the model are strings such as

Now evidently this interpretation presupposes that every such string can be extended by the addition of another "‖." In my earlier discussion (1979–80:156–8) I argued that this is intuitively evident. Both this proposition and the statement that, for any *x, Nx   N(Sx),* obtained from the second introduction rule, are general statements about strings, but they are not obtained by induction. The variable in the second introduction rule is in a way more inclusive than variables over numbers. The thought-experiment that verifies that every string can be extended does not depend on any insight into the specific totality of such strings; this is shown by the fact that in the sense in which a new "‖" can be added to any string, it can be added to any bounded geometric configuration.

Although the generality of the variable *"x"* in the second introduction rule is more inclusive than a variable over numbers (in the present setting, strings), I do not want to say that it ranges over a more inclusive *totality of objects.* There would not be a convincing answer to the question what that totality is. We could specify the range as something like "object given in space or time." But this very general rubric might go with quite different ways of *individuating* such objects. The generality is akin

to the kind of generality Husserl called "formal," characteristic of formal logic. In a sense, we used free variables with a generality interpretation without yet knowing what they ranged over.

If this is our situation, then the impredicativity of induction arises more sharply than before, because it is induction that cashes in our conception of the totality of strings and therefore gives us a predicate that can serve to define the range of variables of quantification that can be used in the formation of predicates.[7] Stated as a general principle, induction is about "all predicates." The predicates that we use are defined by means of basic relations, logical connectives, and quantification over N and other domains that we might come to subsequently. Induction is thus inherently impredicative, because without it we do not have a definite domain for our quantifiers, but we cannot apply it without taking predicates involving quantification over this domain as instances.

It will no doubt occur to the reader that as an "inductive definition" of "N" the introduction rules and the induction principle are incomplete, no matter what resources are deployed for the formation of predicates. Even to obtain the elementary theory of 0 and S, we need the two additional axioms

$$Sx \neq 0$$

$$Sx = Sy \rightarrow x = y$$

which intuitively express the fact that if the successor operation has been applied a number of times beginning with 0, then its application will still yield something *new*. Evidently, since they contain identity, they are bound up with the individuation of the objects of the domain. In the Hilbertian intuitive model, the notion of type provides the principle of individuation (Parsons 1979–80:153–6), and these axioms seem evident enough on intuitive grounds. But other points of view about them are possible, such as Lorenzen's that they are admissible rules once one has introduced the equality predicate in a similar inductive way.[8] These axioms raise many questions, but space does not permit us to go into them here.

We must not take the impredicativity of induction as implying the legitimacy or unavoidability of full second-order logic. Given a domain D for individual variables, full second-order logic is naturally understood by taking second-order variables to range over all subsets of D. It requires that the "predicates" of objects in D should be closed under second-order quantification. Nothing we have said implies this; indeed our picture should suggest rather the opposite, that the "predicates" talked of in the induction rule are a quite open-ended and indefinite totality, depending on linguistic and conceptual resources of whose limits we have no real conception. That would raise a question whether second-order quantifiers, in particular their use to define new predicates as in second-order logic, can have a definite sense. My own view is that any such sense that would license impredicative logic must derive from the concept of set.[9] That such second-order logic is not forced upon us at this point is shown by the fact that a mathematics that assumes the concept of natural number but from there on is strictly predicative is perfectly coherent.

Before going on I want to consider an objection to our thesis suggested, though not in the end endorsed, by Alexander George.[10] He proposes a set-theoretic definition of "natural number".

$$Na =_{df} \forall y[a \in y \land \forall x(Sx \in y \rightarrow x \in y) \rightarrow 0 \in y] \land$$
$$\exists y[a \in y \land \forall x(Sx \in y \rightarrow x \in y)].$$

This definition is a revision of a well-known one due to Quine.[11] The sets in the range of "y" can be assumed to be finite, thus specifiable independently of the class of natural numbers itself. George thus claims that the definition is predicative. George then observes, however, that impredicative instances of the schema of separation will be needed to derive induction for predicates containing quantifiers.

To the claim that the Quinean definition of the natural numbers is predicative, one can also reply that it is so only because the notion of finite set is assumed. George's claim that the above definition is predicative is justified only if "y" is explicitly restricted to finite sets. George does this by a distinctive style of variable for finite sets. To stick to the usual first-order framework, however, I will modify the definition to

$$Na =_{df} \forall y[\text{Fin } y \land a \in y \land \forall x(Sx \in y \rightarrow x \in y) \rightarrow 0 \in y] \land$$
$$\exists y[\text{Fin } y \land a \in y \land \forall x(Sx \in y \rightarrow x \in y)],$$

where "Fin y" means "y is a finite set."

Once one allows oneself the notion of finite set, it seems one should be allowed to use some basic forms of reasoning concerning finite sets. It then becomes possible to reply to George's reply to his objection. The following schema is a natural principle of induction on finite sets:

$$A(\emptyset) \land \forall x \forall y[\text{Fin } x \land A(x) \rightarrow A(x \cup \{y\})] \rightarrow$$
$$\forall x[\text{Fin } x \rightarrow A(x)]$$

If we assume this schema and a parallel principle of recursion, which allows for given terms   and x the introduction of a function symbol f satisfying the equations

$$f(a, \varnothing) = \psi(a)$$
$$f(a, b \cup \{c\}) = \chi(a, b, c)$$

[13]

then by recursion we can introduce the cardinal number of a set:

$$\mathrm{card}(\varnothing) = 0$$
$$\mathrm{card}(b \cup \{c\}) = \mathrm{card}(b) \text{ if } c \in b,$$
$$= S[\mathrm{card}(b)] \text{ if } c \notin b.$$

We can now define the natural numbers in an obvious way:

$$Na =_{\mathrm{df}} \exists x[\mathrm{Fin}\ x\ \wedge\ a = \mathrm{card}(x)],$$

and now induction on natural numbers is straightforwardly provable using induction on finite sets. In this whole development, no set existence assumptions are used except the basic ones giving finite sets: Ø and $b \cup$ *[c]* for given b, c. Given the theory of finite sets, the definition of natural number and the derivation of induction meet any reasonable standard of predicativity.

   As a defense of the claim that induction on natural numbers is after all predicative, this exercise is hardly impressive. What has been assumed about finite sets will just reinforce the reply that although perhaps one can escape the impredicativity of induction on natural numbers, one merely throws the matter back to the notion of finite set, where the same problems will arise. And indeed the assumptions about finite sets used above will clearly give rise to a parallel argument for the impredicativity of Zermelo's induction principle.

   One might seek to escape this conclusion by relying on a settheoretic definition of finite set and a derivation from it of the above induction and recursion schemata. Questions might be raised about the predicativity of these definitions. But in any event the procedure faces the same difficulty as that raised by George about his proposed definition of the natural numbers: for A containing quantifiers, impredicative instances of separation will be needed to derive induction on finite sets.[14]

   There is a positive lesson from our exercise, however. The conclusion that the notion of natural number is predicative relative to that of finite set brings out a relativity of the notion of predicativity, to the conceptual apparatus that is taken as given. This is illustrated further when we consider some discussions of predicativity in the logical literature.

II

Up to now our main argument has consisted essentially in applying to the case of the natural numbers a well-known argument for the impredicativity of inductive definitions. This was applied in particular to so-called "generalized inductive definitions" in which the introduction rules for an inductively defined predicate P may contain general statements involving P. A general form of such definitions consists, for a formula $A(P, x)$ containing the predicate P to be introduced but otherwise only previously understood vocabulary, of the introduction rule

From $A(P, a)$ infer Pa

and the induction principle (elimination rule)

From $[A(\{x: B(x)\}, a) \rightarrow B(a)]$ and Pt infer $B(t)$,

which expresses the idea that P is minimal so as to satisfy the condition given by the introduction rule. Such a means of introducing a new predicate is not always coherent. A sufficient condition is that A should be *monotonic,* that is, that

$$\forall x[A(P, x) \wedge \forall y(Py \rightarrow Qy) \rightarrow A(Q, x)]$$

should hold for all P, Q. I will assume this in what follows.

   If P is introduced in this way, we can think of its extension as built up in stages, indexed by ordinals. At stage 0 the extension P_0 is empty; $P_{+1}a$ holds if and only if $A(P, a)$; for a limit ordinal , P a hold if an only if, for some < , P a holds. Monotonicity implies that, if < , then P a implies P a. If an is reached such that $P_{+1}=P$, then P remains unchanged for all further , and we can take P_a as P. Moreover, assuming the underlying domain to be a set, there will be such a "closure ordinal," for if the closure ordinal has not been reached by stage $a,$ then a must be of cardinality no greater than that of the domain. Thus in set theory, we can prove that monotonic inductive definitions yield welldefined predicates.[15]

   A simple example of such a "definition," relevant to our argument, is that of the notion of accessibility for a relation R of natural numbers, which we will assume to be a linear ordering. Intuitively, "Acc(a)" means that transfinite induction holds for the relation it can therefore be given an obvious second-order explicit definition. In the inductive definition, the introduction rule takes the form

$$\text{From } \forall x[x \, R \, a \rightarrow \text{Acc}(x)] \text{ infer Acc}(a)$$

whereas the elimination rule or induction principle says that, if $A(x)$ satisfies the above closure condition, then it holds for all accessible numbers, that is,

$$\text{From } \forall y[y \, R \, a \rightarrow A(y)] \rightarrow A(a) \text{ and Acc}(t) \text{ infer } A(t).$$

In proof theory, beginning with Gerhard Gentzen's 1936 proof of the consistency of first-order arithmetic, transfinite inductions up to certain countable ordinal numbers have been regularly used.

These ordinals can be represented arithmetically by primitive recursive orderings of the natural numbers. If R is such an ordering, in many cases the notion of accessibility can be used for formal derivations of transfinite induction; standard derivations in texts on proof theory can be recast in the form.[16]

However, one comes rather quickly to need to use instances of the induction schema where the formula $A(a)$ contains "Acc," so that again impredicativity enters. For this kind of reason it has been argued that this and other important generalized inductive definitions are impredicative. For example, in the course of presenting his own analysis of predicativity, Solomon Feferman offers such an argument for the impredicativity of the usual definitions of Kleene's set O of recursive ordinal notations.[17] In the case of Acc, applications of it to prove transfinite induction for shorter well-orderings (such as one of type ε0 that might be used for Gentzen's or another proof of the consistency of arithmetic) can be replaced by other methods that will be recognized as predicative. But this will no longer be true for longer orderings.[18]

The sense of "predicative" in which these last remarks hold is captured by the beautiful and persuasive analysis of predicative provability of Feferman and Schütte.[19] According to this analysis, generalized inductive definitions like that of "Acc" and O are impredicative principles of proof. I shall not enter here into the details of the analysis, which are complex. It turns on looking at transfinite iterations of different methods of enlarging the means of expression and proof of formal systems. The key condition is that if some such method of enlargement is iterated through a transfinite sequence of stages, the stages can be antecedently recognized to be well-ordered. But at no point is any restriction placed on ordinary induction on natural numbers. The considerations advanced above might suggest a doubt as to whether what is predicative according to this analysis really is so. However, in one respect pressing this point would misconstrue the intent of the analysis. What it is intended to capture is predicativity *given the natural numbers,* where the problem of predicativity is raised in the first instance about reasoning about sets of natural numbers, and then perhaps extended to further reasoning about sets and functions. Thus Feferman writes:

> According to this [the predicative conception], only the natural numbers can be regarded as "given" to us.… In contrast, sets are created by man to act as convenient abstractions…from particular conditions or definitions.[20]
>
> (Feferman 1964:1–2)

Feferman and others have suggested that an analysis of predicativity should describe the concepts and principles that are in some way implicit in the conception of natural number. The manner in which reasoning about classes, sets, or functions is allowed is based on the conception of classes as extensions of predicates, and so a generalization about classes (or sets) must be cashable as a generalization about the predicates of which the classes are the extensions. Such generalizations about predicates will be semantical in character, involving satisfaction or truth. Thus semantic reflection, or what Lorenzen some years ago called "logical reflection," is the manner in which second-order entities are understood.[21] By semantic reflection I mean passing from statements using a certain vocabulary to statements applying semantic notions (reference, truth, satisfaction) to that vocabulary. From Tarski's indefinability theorem, it follows that for classical theories semantic reflection is an enlargement of one's means of expression.[22]

One can obtain considerable strength by iterating such reflection into the transfinite, as in the construction of the ramified hierarchy. But on the Feferman-Schütte conception, the iteration is constrained by the requirement that its stages (ordinals) be given in advance as well-founded. In contrast, a generalized inductive definition like that of O or "Acc" allows objects to fall into the extension on the introduced predicate by iteration of the introduction rule (according to the above-described iterative process), but the stages needed, if one does not take them as given from set theory, are described only by the definition itself or another comparable one. This is exactly parallel to the situation with the natural numbers.

The Feferman-Schütte conception of predicative mathematics is constructed from elements that are very basic and deeply entrenched parts of our conceptual apparatus: either first-order logic or a second-order logic with quite minimal comprehension assumptions, the notion of natural number, and semantic reflection. Describing the limits of what can be accomplished by these means not only was a technical achievement but served to delineate a natural conceptual boundary. Some of the discussion of constructivity and predicativity in the immediately preceding period shows the lack of a clear distinction between this apparatus and a more generous conception that would extend it at least by certain generalized inductive definitions. Feferman's criticism of Lorenzen and Wang on this point is justified (Feferman 1964:5). However, I shall argue that the distinction can be seen as one between two senses of predicativity.

If one grants a certain impredicative character to ordinary induction, the issue between Feferman and the earlier writers appears in a somewhat different light. It is hard to see how a mode of concept formation which involves a "vicious circle" in the case of generalized inductive definitions does not involve such a circle in the case of the natural numbers themselves. Granted, then, that Feferman has correctly characterized the limits of predicativity relative to the natural numbers, the case that the traditional arguments deriving from Poincaré show that this is the limit of acceptable mathematics is weakened.

Lorenzen usually characterized his position as "constructivist" or "critical" (in the writings cited in notes 6, 18, and 20) and to that extent did not depend on a particular interpretation of predicativity. However, he did specifically claim that generalized inductive definitions are predicative, though to be sure in a joint paper.[23] Clearly, my view is that this claim is mistaken. However, at this point one should make another distinction. The primary sense of impredicativity applies to sets or classes, and they are said to be impredicatively defined if they are given by abstracts involving quantification over some totality of sets to which they themselves belong. This extends readily to other cases; for example Russell's diagnosis of the semantical paradoxes involves pointing out that sentences such as "The proposition expressed by my present utterance is false" or "Every proposition asserted by a Cretan is false," asserted by Epimenides the Cretan, when taken naively, purport to express propositions that are in the range of their own quantifiers. In the present case, inductive definitions are said to be impredicative because they involve introducing a predicate by rules or axiom schemata such that expressions containing the predicate itself have to be admitted as within their scope. All these cases are of the type where the question of a vicious circle was raised by Poincaré, Russell, and Weyl.

However, if we reflect on the motives of the original critique of impredicativity, an underlying conception was that classes or sets are extensions of predicates and therefore that the circle lay in speaking of sets that could not be extensions of predicates antecedently understood. This conception is quite clear in the writings of Poincaré and Weyl.[24] In the case of Russell it is perhaps not so clear because of his unclear conception of the relation of prepositional functions to language. However, his original adoption of a vicious-circle principle was as a guiding principle in the construction of a "no-class" theory, in which classes were to be eliminable by contextual definition, and propositional functions were treated in a completely predicative way.

Historically, what has served to defuse the critique of impredicativity is set-theoretic realism, with its attendant abandonment of the ideas that sets are extensions of predicates in a given language, so that the domain of sets one can quantify over has to be seen as potential, expanding as one's linguistic resources expand, in particular by quantifying over totalities of sets previously arrived at. Russell took the realistic attitude in a somewhat half-hearted way in introducing his axiom of reducibility, of which he said that it accomplishes "what common sense effects by the admission of classes."[25]

To return to Lorenzen, it is quite clear that he is especially concerned to avoid this set-theoretic realism, what he calls "naive" concepts of set, relation, and function.[26] His concept of set is just the one that underlies the critique of Poincaré and Weyl, even though the issue of a "vicious circle" does not occupy the center of his attention.

Since it is generally agreed that there is a coherent conception of "constructive" mathematics which goes beyond the predicative as characterized by Feferman and incorporates theories of generalized inductive definitions but which still does not presuppose settheoretic realism, it might seem that our discussion can just end with the observation that Lorenzen held such a constructive conception. However, in my view there is still a remark worth making, which justifies to some extent the use of the word "predicative" by Lorenzen and Myhill. Constructivists have not always been very explicit about the relation of higher-order entities to language; indeed Brouwer's own radical view that mathematics is essentially independent of language works against such clarity, particularly in his notion of species (i.e. class). The term "impredicative" was used by Poincaré because of his view of sets or classes as essentially extensions of predicates;[27] in a terminology I have used elsewhere, the language of classes serves as a means of generalizing predicate places in a language. But then that the classes in the range of a generalization should be the extensions of predicates antecedently understood is entirely natural. Now if we call "predicative" such a view of classes, the question arises whether it is violated by inductive definitions. If it is not, it will give a sense in which Lorenzen's view is predicative and a divergence of two possible meanings of predicativity.

Now it is characteristic of the inductive definitions we have been considering that they are introductions of *predicates* and not in the first instance definitions or characterizations of *sets*. Because semantic reflection comes so readily to us, once we have understood such a predicate as the natural number predicate or an accessibility predicate we will almost immediately talk of its extension. However, there is an essential conceptual order here which places the understanding of the predicate before the apprehension of its extension as an object. There is, to be sure, a subtle difference between the situation we are envisaging, where we eschew set-theoretic realism and treat the inductive rules themselves as giving us understanding of the predicate, and a situation where one assumes set theory and where moreover what is being introduced is a predicate of objects in a domain that has been recognized to be a set. In the latter case the axiom of separation implies that there *is* a set that is its extension; there is therefore a proof of its existence which does not use any semantic concept (as indicated above). At least if one is prepared to hold that particular instances of a schema like the axiom of separation can be seen to be true independently of the general principle, it follows that semantic reflection does not enter into one's insight that there is such a set as .

However, this does not change the essential point: that one's understanding of the predicate is prior to the insight that the set exists.

Thus in my opinion Lorenzen does not violate the limitations of his own concept in admitting generalized inductive definitions, and the divergence of two senses of predicativity does indeed exist. Though this observation is a partial defense of Lorenzen against Feferman, clearly the meaning of the term "impredicative" underlying both Feferman's analysis and the earlier part of this paper is so firmly entrenched that it is now the more appropriate way to use the term, particularly since there is no ready alternative term for the same idea.

In these remarks I have bypassed an issue about the status of higher-order entities in constructive mathematics. According to many constructivists, the notion of *function* enters essentially into the interpretations of quantifiers, even over such objects as numbers. The context in which the issue arises is explanation of the meaning of statements in intuitionistic theories in terms of what would count as a proof of them, worked out formally in theories of constructions.[28] In intuitionistic mathematics, we no longer have the mutual reducibility of the notions of set and function which obtains in classical mathematics. A theory may be compatible with the idea of sets and classes as arising by semantic reflection and still depend on a notion of function of a different nature. In my view this issue primarily affects the claim of the sort of conception I have been discussing to be constructive, which has not been a primary concern in the present paper. At all events the conception of function that on this view would be required is weaker than the set-theoretic conception in that the functions that would have to be assumed are intuitively effectively calculable.

Finally, since this paper was originally written for Sidney Morgenbesser, I cannot evade completely a nagging doubt that I am sure would arise in his mind. The model that played the principal role in my argument can be described as one in which induction is constitutive of the meaning of the term "natural number." But, one will ask, is there even a fact of the matter as to what belongs to the meaning of "natural number" as opposed to merely being true (even if necessarily) of all natural numbers? Does not my discussion flirt rather dangerously with the notion of "conceptual truth" or "meaning postulate"?

A proper reply to this objection would require a paper in itself. I will confine myself to some brief remarks. First, even if this model is taken literally it still does not remove every "factual" aspect from the principles of arithmetic: that we *understand* a concept explained in this way and have consistency in its applications is not something we could establish by other more evident or fundamental principles; our possession of a concept of number is a sort of *Faktum der Vernunft*. That state of affairs would obtain even if it were shown that the principles of mathematics are analytic, in the sense of being true by virtue of the meaning of the terms in them.[29]

Second, my argument has taken the form of claims that impredicativity emerges from certain explanations of the meaning of "natural number." In the explanations that I emphasized, induction is directly part of the explanation. But I do not wish to be committed to the thesis that induction is "constitutive of the meaning of 'natural number'" independently of any such explanation. My remarks on some set-theoretic proofs of induction were meant to show that impredicativity will turn up in the context of a variety of such explanations. Still, one is left with the somewhat unsatisfying situation that one has to verify this one explanation at a time, and the possibility exists that someone will turn up a new explanation that will avoid it. Fairly long experience seems to me to make this highly unlikely.[30]

Third, in the present context a purely dialectical reply to the objection is possible: if there is no such fact of the matter in the case of the concept of natural number itself, then on the same grounds there should be no such fact of the matter in the case of notions introduced by generalized inductive definitions. Thus, the thesis of Section II of this paper, that as far as the specific issue of the impredicativity goes ordinary induction is in the same boat as these higher inductions, still stands.

NOTES

1 This observation about proofs using Frege's definition is reinforced by the fact that, in a ramified second- or higher-order logic, if one defines the natural numbers in Frege's way with the second-order quantifier having a definite level or order, then there will be instances of induction of higher level that will not be provable. See Myhill (1974), p. 21. Essentially Myhill's result was stated, with only the briefest indication of a proof, in Wang (1962), p. 642. Kurt Gödel had observed that Russell's attempt to prove the contrary, in Appendix B of the second edition of volume I of *Principia Mathematica* (1925), is fallacious; see Gödel (1944), pp. 145–6.

2 Michael Dummett (1978), p. 199. Dummett does not apply his remark to issues concerning predicativity and inductive definitions in general (see below).

3   This seems to be Dummett's view. Commenting on the result mentioned in note 1, Myhill remarks that if one's "mathematical philosophy" is classical,

> one would reason that the result shows that impredicativity is present in mathematics from the very beginning, i.e. the natural numbers, and that consequently any philosophy of mathematics which repudiates impredicative definitions *ipso facto* repudiates mathematics itself.
> (Myhill 1974:27)

Taken strictly, this claim is very questionable, since a classical view of mathematics hardly requires that the notion of natural number be understood by an explicit definition. If one replaces "impredicative definitions" by the vaguer term "impredicative notions," then an argument for Myhill's claim would require something like the (earlier) observation of Dummett that we are elaborating.

On the other hand Myhill seems to think that the constructivist could avoid this conclusion. The view defended below is that he could do so only at the price of a dogmatic view of the clarity of the notion of natural number and the evidence of mathematical induction.

4   Hilbert and Bernays (1934). The sense in which this model is intuitive is discussed in my "Mathematical intuition" (Parsons 1979–80).

5   As our terminology suggests, the introduction rules and induction can serve as introduction and elimination rules for *"N"* in a naturaldeduction formalization of arithmetic. In such a formulation, the antecedent $A(x)$ in the second premise of induction can be replaced by a premise to be discharged.

6   But see my "The structuralist view of mathematical objects" (Parsons forthcoming).

7   This seems to be the view of Dummett (1978).

8   Lorenzen(1955), p. 134.

9   Parsons (1983), pp. 216–17.

10 George (1987). See also George (1988), pp. 141–3.

11 Quine (1969), pp. 75–7. On some difficulties with Quine's (simpler) version, see George (1987), p. 516, and Section 6 of Parsons (1987).

12 This is a reformulation of an induction principle discovered by Zermelo; see Zermelo (1909). It can be proved in set theory from any one of several definitions of finite set; cf. Levy (1979), p. 77, and Parsons (1987), Sections 3 and 4.

13 The parameter *"a"* can obviously be replaced by a longer, or empty, list of parameters. In a formal theory, we would need to provide for the possibility that the last argument of f is not a finite set, presumably by legislating a throwaway value. In the following application, we assume definition by cases.

14 The same is true of a very interesting set-theoretic proof of induction given by George Boolos (1985), based on an argument due to Dana Scott to prove the axiom of foundation in a theory in which there are variables for partial universes (collections of sets obtained at various "stages" of the iterative conception). Boolos assumes a notion of number, but the assumptions about it are minimal except that numbers serve as stages at which sets are "formed." But for this he assumes a comprehension principle *Spec* (p. 471), and the whole argument turns on the use of the notion of groundedness (p. 472), which makes the use of impredicative instances of *Spec* unavoidable.

Boolos is not concerned with the question of predicativity but rather with the claim that his assumptions are "significantly less inductive in character" than mathematical induction or other principles that have been used to derive it. The significance of Scott's argument and Boolos's use of it certainly deserves further discussion.

Scott's original presentation is in "Axiomatizing set theory" (1974). See also Shoenfield (1977).

15 For information concerning generalized inductive definitions and their importance for proof theory see Feferman and Sieg (1981). Cf. also Aczel (1977). This informative article is not as introductory as its title would suggest.

16 For example, this is true of the derivations in Section 21 of Schütte (1977), which are predicative by the Feferman-Schütte criterion discussed below, and of those in Section 22 of Schütte (1960), which are not. On the latter, cf. his "Logische Abgrenzungen des Transfiniten" (1962), p. 110. To handle the ordering dealt with in Section 29 of *Proof Theory* (1977), one needs arbitrary finite iteration of inductive definitions.

All these orderings are primitive recursive, and their elementary properties can be proved in primitive recursive arithmetic. It is in the problem of deriving transfinite induction that the constraint of Gödel's Second Incompleteness Theorem shows itself. Gentzen's proof, for example, can be carried out with a single instance of induction on an ordering of order type ϵ_0. Thus this instance cannot be derived in first-order arithmetic.

17 Feferman (1964), p. 5.

18 If the inductive definition of accessibility is our only means of proof beyond arithmetic, we need to use the elimination rule with predicates containing "Acc" even in some cases that are clearly predicative given the natural numbers. Does this cast doubt on our argument for the impredicativity of ordinary induction? In the present situation, such *prima facie* impredicativities can be replaced by other methods of proof. For example, one can use ramified second-order logic (i.e. second-order logic with the second-order variables assigned levels, so that the variables of a given level can be interpreted to range only over second-order entities defined by means of quantifiers of lower levels), where, if transfinite levels are needed, they can be antecedently shown to be well-founded. No

such alternatives are in sight in the case of ordinary induction. Nonetheless, some further logical analysis to reinforce this point would be desirable.

19 Feferman (1964). This is still the best source for his motivating ideas and for the statement of the most basic technical results, although his analysis is refined and extended in later papers.

   Schütte offered an analysis based on the same basic ideas and obtained independently some of the relevant technical results but did not carry the matter as far as Feferman, mainly because the only form of predicative analysis that he considered was based on the ramified hierarchy. In "Logische Abgrenzungen" (1962) he presents his ideas very clearly and describes his results informally.

20 Schütte is not so explicit on this point.

21 Lorenzen (1958), p. 244. In recent work Feferman has explored the idea of "reflective closure" of theories, to capture the idea of "closing" a theory under semantic reflection. Predicative analysis as he has previously described it turns out to be what he calls the "strong reflective closure" of first-order arithmetic. See his forthcoming "Reflecting on incompleteness."

22 Semantic reflection and its role in mathematics are discussed in considerable detail in *Mathematics in Philosophy* (Parsons 1983), particularly Essays 1, 3, 8, and 9.

23 Lorenzen and Myhill (1959), pp. 47–8. Hao Wang also proposes that generalized inductive definitions be included in predicative theories, for example, in Wang (1962), Section 5. On p. 644, there is intimation of a *ceteris paribus* argument like that of the present note.

24 Though I find it clear enough, in Poincaré it is not quite so explicit, or so clearly disengaged from other considerations such as rejection of the actual infinite, perhaps because of his negative attitude towards symbolic logic. But he makes quite clear that he expects sets to be definable, most explicitly in the first essay of *Dernières Pensées* (1913); see especially the criticism of Cantor's diagonal argument in Section 6.

   For Weyl, see the analysis of the concept of set in *Das Kontinuum* (1918), Section 5, and its polemical use in "Der *circulus vitiosus* in der heutigen Begründung der Analysis" (1919).

25 Russell (1956), p. 81.

26 Lorenzen (1958), pp. 246–7.

27 The first version of this paper said that the term was "coined" by Poincaré. George (1987) pointed out that Poincaré credits the introduction of the term to Russell (p. 514).

28 One sees this clearly in W.W.Tait's characterization of the mathematics that can be obtained *without* presupposing a general concept of function. See Tait (1981). The most elaborate theory of the kind I am alluding to is Per Martin-Löf s intuitionistic theory of types; see Martin-Löf(1984).

29 Cf. Kurt Gödel's rather cryptic discussion of the analyticity of the axioms of *Principia* near the end of "Russell's mathematical logic" (Gödel 1944). The connection between this possibility and Quinean criticisms of analyticity is commented on briefly in my introductory note to the paper in *Collected Works,* vol. II, pp. 115–17. Similar issues are discussed interestingly in a larger context in Wang (1985).

30 We also have to remember, as Hilary Putnam reminded me, that arithmetic was part of mathematics for many centuries before the principle of mathematical induction was formulated. (It apparently dates from the seventeenth century.) This consideration raises a number of questions that I cannot deal with here. One would have to investigate explanations of the notion of natural number that were given in the earlier times and inquire how clearly they differentiated the *natural* numbers from other number systems. That such an explanation was ever given that would meet later standards of adequacy is very doubtful.

REFERENCES

Aczel, P. (1977) "An introduction to inductive definitions," in J.Barwisc (ed.) *Handbook of Mathematical Logic,* Amsterdam: North-Holland, 739–82.

Boolos, G. (1985) "The justification of mathematical induction," in *PSA 1984,* vol. 2, East Lansing, MI: Philosophy of Science Association, 469–75.

Dummett, M. (1978) "The philosophical significance of Gödel's theorem," in *Truth and Other Enigmas,* London: Duckworth, 186–201 (first published 1963).

Feferman, S. (1964) "Systems of predicative analysis," *Journal of Symbolic Logic* 30:1–30.

——(forthcoming) "Reflecting on incompleteness."

Feferman, S. and Sieg, W. (1981) "Introduction," in W.Buchholz, S. Feferman, W.Pohlers and W.Sieg (eds) *Iterated Inductive Definitions and Subsystems of Analysis* (Lecture Notes in Mathematics 897), Berlin: Springer.

George, A. (1987) "The imprecision of impredicativity," *Mind* 96:514– 18.

——(1988) "The conveyability of intuitionism, an essay on mathematical cognition," *Journal of Philosophical Logic* 17': 133–56.

Gödel, K. (1944) "Russell's mathematical logic," in P.A.Schilpp (ed.) *The Philosophy of Bertrand Russell,* Evanston, IL: Northwestern University Press, 123–53. Reprinted in S.Feferman *et al.* (eds) *Collected Works,* vol. 2, *Publications 1938–74,* New York: Oxford University Press, 1990.

Hilbert, D. and Bernays, P. (1934) *Grundlagen der Mathematik* I, Berlin: Springer (2nd edn 1968), Section 2.

Levy, A. (1979) *Basic Set Theory,* Berlin: Springer.

Lorenzen, P. (1955) *Einfurung in die operative Logik and Mathematik,* Berlin: Springer.

——(1958) "Logical reflection and formalism," *Journal of Symbolic Logic* 23:241–9.

Lorenzen, P. and Myhill, J. (1959) "Constructive definition of certain analytic sets of numbers," *Journal of Symbolic Logic* 24:37–49.

Martin-Löf, P. (1984) *Intuitionistic Type Theory,* Naples: Bibliopilis.

Myhill, J. (1974) "The undefinability of the set of natural numbers in the ramified *Principia,"* in G.Nakhnikian (ed.) *Bertrand Russell's Philosophy,* New York: Barnes & Noble, 19–27.

Parsons, C. (1979–80) "Mathematical intuition," *Proceedings of the Aristotelian Society,* NS 80:145–68.

——(1983) "Sets and classes," reprinted in *Mathematics in Philosophy,* Ithaca, NY, and London: Cornell University Press (first published 1974).

——(1987) "Developing arithmetic in set theory without infinity: some historical remarks," *History and Philosophy of Logic* 8:201–13.

——(forthcoming) "The structuralist view of mathematical objects," *Synthese.*

Poincaré, J.H. (1913) *Dernières Pensées,* Paris: Flammarion (2nd edn 1926).

Quine, W. (1969) *Set Theory and its Logic,* revised edn, Cambridge, MA: Harvard University Press.

Russell, B. (1925) *Principia Mathematica,* 2nd edn, vol. 1, Cambridge: Cambridge University Press.

——(1956) "Mathematical logic as based on the theory of types," in *Logic and Knowledge,* London: Allen & Unwin.

Schütte, K. (1960) *Beweistheorie,* Berlin: Springer.

——(1962) "Logische Abgrenzungen des Transfiniten," in M.Kasbauer and F. von Kutschera (eds) *Logic und Logikkalkul,* Freiburg and Munchen: Alber, 105–14.

——(1977) *Proof Theory,* Berlin: Springer.

Scott, D. (1974) "Axiomatizing set theory," in T.J.Jech (ed.) *Axiomatic Set Theory,* Proceedings of Symposia in Pure Mathematics, vol. 13, part 2, Providence, RI: American Mathematical Society, 207–14.

Shoenfield, J.R. (1977) "Axioms of set theory," in J.Barwise (ed.) *Handbook of Mathematical Logic,* Amsterdam: North-Holland, 321–44.

Tait, W.W. (1981) "Finitism," *Journal of Philosophy* 78:524–46.

Wang, H. (1962) "Ordinal numbers and predicative set theory," in *A Survey of Mathematical Logic,* Peking: Science Press (first published 1959). Reprinted Amsterdam: North-Holland, 1963.

——(1985) "Two commandments of analytic empiricism," *Journal of Philosophy* 82:449–62.

Weyl, H. (1918) *Das Kontinuum,* Leipzig: Veit.

——(1919) "Der *circulus vitiosus* in der heutigen Begrundung der Analysis," *Jahresbericht der Deutschen Mathematiker- Vereinigung* 28: 85–92.

Zermelo, E. (1909) "Sur les ensembles finis et le principe de 1'induction complete," *Acta Mathematica* 32:183–93.

THREE INSUFFICIENTLY ATTENDED TO ASPECTS OF MOST MATHEMATICAL PROOFS: PHENOMENOLOGICAL STUDIES

Robert S.Tragesser

"Baroque" is the name of one of the forms of the syllogism; the eighteenth century applied it to certain excesses in the architecture of the century before.

Borges

We do not learn to demonstrate from the manuals of logic, but from the books which are full of demonstrations, which are the mathematical and not the logical.

Galileo

SUMMARY

This essay is primarily concerned to make the following points:

1 that formal-logical structure is not essential to mathematical proof and, at times, can even serve to conceal the important role that *understanding* plays in it,

2 that a well-described possible proof can have the same epistemic benefits as an actual proof (thus showing that a clear line is to be drawn between mathematical and empirical warrant and providing a promising direction to pursue in trying to explain the *a priori* character of mathematical knowledge),

and

3 that a proof or chain of proofs cannot leave anything unproved, contrary to the common idea that proofs must begin with *assumptions* that are not themselves proven.

These facts point to difficulties in regarding proofs either as themselves being formal-logical derivations or as being satisfactorily represented by them. The formal-logical model idealizes away aspects of proof that are vital to mathematical thought, particularly obscuring the complex role that understanding plays in it.

In their more important parts the *phenomenological* studies to follow seek to point out the following.

First, formal-logical rules can be significantly superfluous to mathematical proof. This is contrary to the idea many have—philosophers rather than mathematicians—that proofs just are constituted by steps appealing to such rules. It will be seen that the steps of many (indeed, virtually all) proofs which are to be found around and about in the world cannot be regarded as following, as being structured by, the laws of familiar logical calculi (even in the supposedly "formalist" tract, David Hilbert's *Foundations of Geometry* (1971)), but are nevertheless clearly valid, as clearly valid usually as steps in logical calculi governed by formal-logical syntactical rules, perhaps even more clearly valid. This tends to confirm, say, John Locke's view about the superfluousness of "syllogistic"—although the formal-logical has its uses (as Leibniz would have insisted), principally today in metalogical/metamathematical investigations (which may themselves lead to mathematical invention, as in the creation of nonstandard analysis) as well as in computer science. But, most importantly, the representation of proofs by means of formal calculi conceals the deep role that *understanding* plays in proof. Any adequate philosophy of mathematics must come to grips with this role of understanding. Outside of mathematics, this phenomenon of most mathematical proofs not following in outline the structure that would be required by familiar formal-logical calculi also suggests that those who too easily read a logical syntax into ordinary language ought well to have second thoughts.

Second, our studies will strongly suggest that a (well-described) *possible* proof will do for an actual proof in that to have had sufficient understanding of a possible proof in order to see that it is a possible proof of a theorem is, in effect, to find that the theorem considered has a proof. This as far as I know is a hitherto unnoticed trait of proofs.

1 It sharply distinguishes proof as a warrant from empirical evidence as a warrant—to contemplate a (well-described) possible empirical warrant (such as an experiment and its results) and to find that, if such occurred, it would indeed be a warrant for such and such an assertion, will not lead to the conviction that the assertion has a warrant (the experiment has to be performed, the results thereby secured).

2 That trait of possible proofs enables us to give a rough but highly promising idea for a characterization of mathematics—mathematical problems are problems that can be solved "in one's head." Proofs have the trait of possibility sufficing for actuality by virtue of their being artifices of the understanding— that such artifices are available for mathematical problems is what enables the problems to be solved "in one's head." This peculiarity of mathematical problems obviously enables us to elaborate a sense in which mathematical truths are known *a priori* (because proofs are artifices of the understanding), but also the sense in which sound mathematics yields *necessary* truths—since possible proof in a sense suffices for actual proof, if there is a possible proof for *s,* then there is no possible proof for Negs.

Third, proofs or chains of proof can leave nothing unproved. This is contrary to the idea that proofs must begin with assumptions not themselves proven. But, for example, Hardy and Wright's *An Introduction to the Theory of Numbers* (1979) makes no mention of axioms, of unproven assumptions, and yet one does not have the sense that the proofs are somehow logically incomplete by virtue of unproven assumptions not having been set out as such. It is understanding that stands in for unproven assumptions. It is pointed out that reasoning, in arithmetic, centering on assertions "take the least number such that…" is justified on the basis of the understanding of the system of numerals; the very fact that the logical analysis of such assertions leads to second-order assertions shows that the logical account of the least number principle does not give an adequate representation of our understanding, our understanding on which arithmetic, number theoretic, reasoning is based, and so must be understood as based on our intuitive understanding of our numerals!

Those three traits of proof, elaborated below, make it difficult to regard proofs, as many do regard proofs, and as I have remarked, either as just derivations in formal systems or as satisfactorily supposed to be such derivations. The conception of proofs as formal derivations enabled David Hilbert (1965) to implement his idea of having a mathematical theory of proof (on a par, as he put it, with the philosopher's philosophical theory of reason or the physicist's physical theory of experimental apparatus). It also permits—not to their benefit—philosophers of mathematics principally interested in ontology to ignore proof, treating of formallogical systems instead. As to the ontologists, since proof is the principal means by which the mathematician comes to the truth, it would seem that proof as it occurs in living or written mathematics should be of intense interest for in proofs could be found what mathematics presupposes about its objects. But more importantly than all of this, as I have also remarked, in fact such idealization of proofs as formal-logical derivation, as the traits indicated above and developed below reveal, idealizes away aspects of proofs vital to mathematical thought, aspects giving deep clues to its nature — most especially the deep, complex, and multifarious role of understanding in mathematical thought.

The idea of phenomenological work in philosophy

The idea of phenomenological study is notoriously difficult to articulate in any refined way, and I shall not attempt this. Crudely, the idea in the case of mathematical proofs is to go back to the things themselves, to actual proofs, and think them through, look at them this way and that, sound them out, try to see what makes them tick—rather than proceeding through some idealization or theoretical representation of proof. On my view, Edmund Husserl's "transcendental phenomenology"[1] was an attempt to provide grounds for the conviction that what might be learned in this way would have objective content, but I am only interested in the practice of phenomenology, in the way that Edmund Husserl's *Logical Investigations* (1913) was only interested in practicing phenomenology.[2] It is doubtless so that what can be managed by way of phenomenological studies will have conjectural and idealizing elements (will be fallible), but it is part of the phenomenological attitude to evade these as much as one can (and it takes a lot of work to learn how to evade them). In any case, the phenomenological way of proceeding in philosophy is the very opposite of that way of proceeding espoused by, for example, David Lewis (1986) justifying a metaphysical claim—"Because the hypothesis is serviceable, and that is reason to think that it is true…. "Further discussion of the virtues of doing philosophy on a basis of phenomenological studies and the vices of doing it on the basis of "serviceable hypotheses"—the pseudo-scientific methodology—will be found *inter alia* below. The main thing is that the phenomenological attitude is one of craving genuine understanding and of not relenting in this quest, of not settling for merely serviceable hypotheses, until all roads that could possibly lead to better understanding have been followed down to their very ends.

I.
HOW MATHEMATICIANS GET BY (BETTER) WITHOUT FORMAL-LOGICAL RULES

Formal-logical rules, such as logical rules of inference, can be significantly superfluous to mathematical proof. Consider, as we shall, the opening development in Hilbert's *Foundations of Geometry* (editions from 1899). Despite its almost successful attempt to set down adequate and adequately stated axioms for the development of Euclidean geometry and, above all, its emphasis on what we would now call metamathematical or metalogical themes (e.g. independence, consistency), Hilbert does not set out a logical syntax and, *a fortiori,* does not give formal-logical rules, laws. In his textbook with Wilhelm Ackermann

Principles of Mathematical [Theoretical] *Logic (Grundzüge der Theoretischen Logik)* (editions from 1928), where logical syntaxes and logical rules and rules of inference are set out and studied, a distinction is made between purely intuitive logical thinking and logical thinking reflected and conducted in a logical calculus (p. 1), such as they provide, a calculus in some strong sense adequate to the needs of mathematical demonstration, unlike the Aristotelian syllogistic. It was, of course, Gottlob Frege (in work from 1879 through 1893) who made possible the writing out of mathematical demonstrations by a wholly (as he said) logically "gapless method," and it was, among others, Giuseppe Peano (from *circa* 1889) who made this feasible (by, among other things, a better notation)—made feasible the presentation of mathematical disciplines/theories in terms of a logical calculus. In any case, Hilbert's *Foundations of Geometry* contains "purely intuitive logical thinking"— thinking unembedded in a logical calculus. And it is not at all as if, but for a proper symbolization and mentioning rules of inference, say, the *Foundations* is in effect or at just one remove from being written out in a logical calculus—a close look at the proofs there show that things stand quite otherwise, and this is generally how it is with mathematical demonstrations, even when they make a pretense of being formal-axiomatic. The question, then, is, how does intuitive logical thinking manage, in the main, to sustain valid logical inferences? If one were satisfied with a philosophical methodology such as espoused by David Lewis (quoted above), then perhaps one might, as for example Gilbert Harman (1973) did, simply propose that our brains, at an unconscious level, just are performing the syntactical operations. But the attitude of those to whom phenomenological work in philosophy appeals is that just having serviceable hypotheses is rankly unsatisfying when a better, more direct, and genuine understanding might be available. That is, we want a more direct and satisfying answer to the question of how "intuitive logical thinking"—rather than demonstration via a logical calculus—works, ticks, the answer of Harman *et al.,* that we are really thinking in a logical calculus at the unconscious level of the computation or computer language of our brains being quite unsatisfying, unsatisfying because not based on anything like insightful understanding secured from the study of the generally formal-logic free proofs in mathematics. In any case, in order to give us a wider perspective on the problem of how we manage to make logically valid inferences without appealing to or otherwise explicitly utilizing logical rules of inference, a little more history is useful before our phenomenological work proper.

"I think," wrote John Locke,

> everyone will perceive in mathematical Demonstrations, that Knowledge gained thereby, comes shortest and clearest without Syllogism.
>
> (Locke 1975:672)

Locke not only held that in mathematical demonstrations knowledge comes shortest and clearest without syllogism (as everyone can see), but, further,

> Syllogism serves our Reason [in that it shows] the connexion of [intermediating ideas between premisses and conclusion] in any one instance and no more; but in this, it is of no great use, since the Mind can perceive such connexion where it really is, as easily, nay, perhaps better without [Syllogism].
>   If we observe the Actings of our own Minds, we shall find, that we reason best and clearest, when we only observe the connexion of the [ideas], without reducing our Thoughts to any Rule of Syllogism.
>
> (Locke 1975:670)

As was said, Hilbert distinguished between "purely intuitive logical thinking" and "logical thinking [as] reflected in a logical calculus." Indeed, as also remarked, Hilbert's own *Foundations of Geometry* does not give either logical rules of inference or any other sort of logical laws. That is, it develops geometry by means of "purely intuitive logical thinking." Logical thinking as reflected in a logical calculus would proceed via rules of inference and logical laws (as well as a logical syntax) given beforehand. The difference is plain. In the case of intuitive logical thinking one somehow *sees that* (although perhaps in some complex way) one thing follows from another without appealing to rules of inference; but in the case of thinking via a logical calculus one sees that given one thing, and then successively applying the logical rules, such and such other things calculatively result. It is an *interpretation* of those rules that one thing will thereby logically follow from the other, but one need not understand this in order to see that the rules have been followed. In order to be able to find that one thing results from applying logical operations/rules to another one need not see that one thing is thereby validly deduced from those others. What we are interested in is the phenomenon of finding that one thing is validly inferred from others without seeing that the one thing calculatively results from applying valid, formal-logical rules of inference (anything but, *pace* Harman's claim that we are performing such calculations whether we know it or not).

Leibniz (1981:479) gave us a useful concept here, that of "formal." The principal trait of a *formal* argument is that *the form of reasoning has been demonstrated in advance* as being correct. He gives as examples not only rules and fixed complexes of syllogism, but also formal algebraic reasoning, such as the use of the justified quadratic formula, or formal reasoning in infinitesimal analysis, such as the justified law that the differential of a sum is the sum of the differentials (p. 479). An

important trait of such reasoning is that it can proceed by the forms alone, "with no need of anything to be added." Thus we might distinguish between formal and informal reasoning and, specifically, between formallogical and informal-logical reasoning. Formal-logical reasoning corresponds to reasoning Hilbert characterized as reasoning reflected in a logical calculus, while informal-logical reasoning corresponds to what Hilbert called "purely intuitive logical thinking." Notice that formal reasoning includes reasoning with, for example, proven correct algebraic formulas, such as the quadratic formula for calculating the roots of quadratic equations. The term "logical" in "formal-logical" and "informal-logical" carries the specific connotation that the argument turns on something verbal, on verbal representation. Thus a formal-logical form is crucially structured by, or is otherwise supervenient on, linguistic form, although this need not be true of "the formal" generally, as in algebra or infinitesimal analysis.

Locke is claiming that, as anyone can see, in mathematical demonstrations in which knowledge comes shortest and clearest, such knowledge will come by *informal* reasoning (and this is not to say that *all* knowledge coming by informal reasoning will be knowledge that thereby comes shortest and clearest). Just antecedent to those remarks Locke observed,

> If Syllogism must be taken for the only proper instrument of reason and means of Knowledge, it will follow, that before Aristotle there was not one Man that did or could know any thing by Reason; and that since the invention of Syllogisms, there is not one of Ten Thousand that doth.
>
>   But God has not been so sparing to Men to make them barely two-legged Creatures, and left it to Aristotle to make them Rational…. God has been more bountiful to Mankind than so. He has given them a Mind that can reason without being instructed in the Methods of Syllogizing. The Understanding is not taught to reason by these Rules; it has the native Faculty to perceive the Coherence, or Incoherence of its Ideas I say not this any way to lessen Aristotle…who did in this very invention of Forms of Argumentation, wherein the Conclusion may be shown to be rightly inferred, did great service against those, who were not ashamed to deny any thing. And I readily own, that all right reasoning may be reduced to his Forms of Syllogism. But yet I think without any diminution to him I may truly say, that they are not only, nor the best way of reasoning, for the leading of those into Truth who are willing to find it, and desire to make the best use of their Reason, for the attainment of knowledge.
>
> (Locke 1975:671–2)

There is no doubt that Locke is right—we reason well, and ubiquitously men reason, without appeal to rules of inference. Leibniz seems to agree with Locke in this and corrects and in a sense reinforces Locke by pointing to logical inferences that could not be represented in terms of syllogistic reasoning that was then available (in contrast with what is available *post* Frege). Leibniz does not follow Locke in his assessment of the low value of formal and formal-logical reasoning, and Locke thinks that all right reasoning may be reconstructed syllogistically. The point is that both Locke and Leibniz recognize that inferential reasoning can be carried out on the basis of understanding without appeal to formallogical laws; and, to repeat, Leibniz's example of valid inferences that cannot be reconstructed in then known syllogistic shows in a very particular way that Leibniz admitted inference that was based on understanding rather than formal-logical rules—*presurhably the understanding that would as it were prove the formal-logical laws,* recalling that for Leibniz (and this is the usage we have adopted here) something *formal* is a form of reasoning that has been "demonstrated in advance" as being correct.

So the suggestion emerges that there is a strong distinction to be made between formal-logical and informal-logical reasoning, wherein the former appeals to logical laws and the latter makes no appeal to laws of logic whatsoever, not even covertly, *so that such reasoning is misrepresented if formal-logically represented.* It is possible to reason logically without utilizing a formal logic indeed, as several times remarked, it is what we and, in particular, mathematicians do almost all the time. *But how do we manage this, and what is the relation between informal-logical reasoning and formal-logical reasoning?* Certainly our consideration of LockeLeibniz might suggest that the understanding on which informal reasoning trades is the or a basis for the choice of formal-logical rules, the basis for "demonstrating" those rules in advance, for convincing ourselves of their correctness, a kind of convincing that of course does not itself require (circularly) rules of inference, but understanding only.

Locke and Leibniz gave examples of demonstrations proceeding without appeal to any syllogism, to any rule of inference.[3] But examples from David Hilbert's highly deductive but formal-logic free *Foundations of Geometry* will serve us better. It would be especially valuable for us to have examples of arguments that turn heavily on logical words or concepts but definitely without appeal to formal-logical laws.

One of Hilbert's axioms is (I.3):

  There exist at least three points that do not lie on a line.

Now one of the things I want to make a point of is that, as Locke and Leibniz recognized, *the understanding can stand in for formal logic* and, indeed, that the occasions on which it does this can sometimes (maybe frequently) form grounds for

justifying insightfully certain logical laws, so I want to dwell a moment on the statement of this axiom, using it as a first occasion for showing how the understanding can indeed stand in for logic.

Notice that anyone who is acutely logic-minded will find this axiom highly, maddeningly, polysemic, while someone who is geometry-minded will not. The logic-minded will ask, does this mean that there are three points, none of which lie on a line at all, or does it mean that there are three points such that, for any given line, not all three points lie on the given line, or could it even mean that there is one line on which three or more points do not lie?

Had the axiom been expressed in a familiar symbolism of quantificational logic, there would not have been that ambiguity. But someone geometry-minded would not have especially noticed the ambiguity since they would have it in mind that it could not happen that two points were not on a line. Here is a case where understanding keeps one from finding ambiguity in statements which form a logical point of view are ambiguous. That is, understanding stands in for, and serves part of the purpose of, formallogical precision. It is perhaps worth remarking that there is a familiar experience of a very logic-minded person without much mathematical understanding coming to mathematical texts and finding them indeed maddeningly polysemic or confused, whereas they are not so to the mathematician because understanding governs the reading and writing of the texts more than the literal on which the logic-minded tend to focus. (I remember very vividly working with two mathematicians, and one was confounding me until I discovered that his statement of a theorem, if logically analyzed, would have a universal qualifier followed by an existential quantifier, but he was clearly *proving* a theorem which he believed himself to have stated that had the quantifiers the other way around. I said some version of, "You mean 'there is/for all', not 'for all/there is,'" and he replied, rather more indifferently than sourly, *"whatever,"* and the other mathematician was not in the least appalled. The other mathematician in effect said to me, don't be a dope, he *couldn't possibly* have meant "for all/there is." The misstatement stayed on the blackboard for at least three weeks.) But it would be wrong, or not at all necessarily true, to say anyone but I had the correct *statement* somewhat in mind. Hilbert, for example, might never have stated that ambiguous axiom unambiguously. *Understanding* keeps things under control.

But let us now consider the theorem from Hilbert's *Foundations* together with a restatement of the axiom in the direction of greater logical informality—the greater the deviance from the routines of the usual formal quantificational logic, the more we will be able to see how, in the spirit of Locke, we could cogently reason before Aristotle or, for us, Gottlob Frege. *Warning:* It is extremely important that one distinguish sharply between the phenomenologically presented informal reasoning and its highly reconstructed formal-logical representation, that one not suppose that, in the presence of the former, one is also in the presence of the latter.

Axiom I.3: At least three points do not lie on a line.

Theorem A: At least one point is not on a given line.

I want to look at three proofs or demonstrations of this theorem from that axiom. The first two are highly informal, though the second is more elaborate. The third is as nearly formal as I have patience to give.

Proof I: Otherwise all points would be on that line, contradicting the axiom that at least three points are not on a line.

Proof II: The only way the theorem could fail us is if all points were on some given line. But if all points were on one line, it could not be that some three points are not on a line, contradicting the axiom.

In order to give the more formal proof, we need a more formal statement of the theorem,

Theorem A:.

Proof III:

1 Suppose.

2 $\exists L \neg \exists x [\neg x \text{ on } L]$.

3 $\exists L \forall x \neg [\neg x \text{ on } L]$.

4 $\exists L \forall x [x \text{ on } L]$.

5 $\exists x_1 \exists x_2 \exists x_3 \forall L [x_1 \neq x_2 \neq x_3 \neq x_1 \ \& \ [\neg x_1 \text{ on } L \lor \neg x_2 \text{ on } L \lor \neg x_3 \text{ on } L]]$ (Axiom).

6 $[x_1 \neq x_2 \neq x_3 \neq x_1 \ \& \ [\neg x_1 \text{ on } L \lor \neg x_2 \text{ on } L \lor \neg x_3 \text{ on } L]]$.

7 $[\neg x_1 \text{ on } L \lor \neg x_2 \text{ on } L \lor \neg x_3 \text{ on } L]$.

    From 4:

8 $x_1 \text{ on } L, x_2 \text{ on } L, x_3 \text{ on } L$.

But this contradicts 7.

Each step of this (almost) formal-logical demonstration can be regarded as being determined and justified according to a (to us, now, familiar) formal-logical law specified in advance, for example, "*not for all* is *some not*" or "*not-not-p* is *p*." In that case the only understanding required for checking or following the demonstration is that such a rule applies at each step.

However, as Locke rather vigorously observed, one can think through the steps 1–8 informally, seeing that each step is justified independently of, wholly ignoring, the rules—as perhaps the reader has done! Thus, I can find it obvious or evident that *"not all"* just is the same thing as *"some not."* I find this by virtue of my understanding of *"not all"* (I will discuss below

the matter of how one might not have just this understanding). This is *the same sort of understanding* that enabled me to find in the case of, say, Proof I that *"The only way the theorem could fail is if all the points were on the given line,"* except that in this case there is no familiar formal-logical law that directly covers the inference (a matter— one central to our concerns here —also to be discussed below). It is by that sort of understanding that both Proof I and Proof II go through, are convincing. Gottlob Frege, in his *The Basic Laws of Arithmetic* (1967), provides an expression of the informal understanding justifying considered formal-logical laws. For example, for *modus ponens:*

> [MP] "from **A** and **A** → **B**, infer **B**" is valid since if **B** were false and **A** true, **A** → **B** would be false, but **A** → **B** is given as true.

Such a use of understanding is not in any awkward or circular sense a use of logic to justify logic,[4] but a use of understanding to demonstrate or justify formal-logical laws, for the justification of formal-logical laws may itself, even though reasoned, nevertheless not draw on formal-logical laws, but, rather, on the understanding. It might be thought that the argument [MP], suitably unpacked or expanded, involves an instance of *modus ponens,* that it uses the very rules it means to justify. But this is wrong. It is a deep confusion to suppose that our understanding were exploiting formal-logical laws to justify such. The priority of understanding to formal-logical rules is, as I have suggested, evident in the case of Proofs I and II which do not manifest the pattern of demonstrations of familiar formal-logical laws or of laws we are readily prepared to formulate.

In the case of the near formal-logical Proof III we can indeed so deploy our understanding that we do not need to appeal to formallogical laws. Indeed, here we can regard this understanding as justifying those rules. It might be thought that no such justification of a general rule would be forthcoming if what is being justified is a step involving one or two particular propositions. But one easily secures generality by noticing that the justification of the particular logical operation does not depend on the other parts of the syntax of the considered proposition.

When we ask what the understanding is an understanding of, perhaps the best we can say is, for example, the understanding is an understanding of *"not all,"* resisting going on to say something like "words" or "meanings." That there is such understanding is undeniable. Crudely: otherwise there would not have been any man that reasoned before Aristotle/Frege. Especially one could not be compelled by Proofs I and II which are quite estranged from any pattern fitting familiar formal-logical laws. And yet we, and at least mathematicians, make such inferences hourly. Even if it is true that explanations must come to an end somewhere and we have to say, we *just understand,* still, in the case of, say, "*not for all* is the same as *some not,*" the understanding is not above criticism, not immune to refinement and bolstering.

For example, a Brouwer/Heyting intuitionist/constructivist would claim such an "inference" invalid because you could in general end up asserting the existence of something for which the proof of the assertion provides no means for giving in any more definite a way. But we can notice by looking at the intentions which control the use of *"not all"* in the above context that one's attitude is that all the points are there, all the lines are there, and all relations between them are already settled or fixed or cast. It is by pointing out this attitude that one points to an important aspect of one's understanding by virtue of which "*not for all* is the same as *some not*" holds. If one realized that the domain over which quantification was being performed was somehow unfinished or in growth or otherwise changing, then an inference based on an understanding of *"for all"* etc., that yielded that "*not for all* is the same as *some not*" would not be valid, quantification taken in that sense would therefore be *inappropriate*. I do not mean to make light of a subtle and difficult issue. The point here is just that the use or meaning of an expression for logical operations does not compel us to accept inferences made according to rules based on the understanding of that sense as valid. Rather, we are free to "project" (Wittgenstein's word) and vary the sense of logical operations according as other reflections (such as insight into the nature of the domain being reasoned about) demand.

The question arises, by virtue of what understanding are Proofs I and II taken to be cogent, sound, valid? If we look at how those proofs differ from the more nearly formal Proof III, we find that the complexities of quantifiers that get into the formal proofs are evaded in the first two informal proofs (I and II), and thus if we look for what enables us to evade those quantifier constructions, we will find a way in which understanding is standing in for formal-logical rules of inference. In the informal proof the cogency of the considerations rests on the understanding of the expressions "at least one," "at least three," "a given" which work to enable one to evade the classical quantificational structure. One could perhaps formulate a syntax and formal-logical laws that could be used to give those proofs a more faithful reconstruction or representation than Proof III, *but one would expect that we could not generally use these now formalized devices to serve all the purposes classical quantificational structure serve.* It might be conjectured that Proof II, while valid, is not so much less rigorous than the more formal Proof III as it is shorter because it evades the more general logical structure formal, classical quantificational logic brings with it. Locke, I think, was right—*the informal proofs are generally shorter because we can generally manage to substitute less general, weaker, logical devices for those for which we have nice formal-logical laws.* I do not know how to argue for this before we have at hand the results of a delicate and wide-ranging study of such informal logic devices as those involved in Proofs I and II. But I will let this stand as one conjecture as to why informal proofs are shorter.
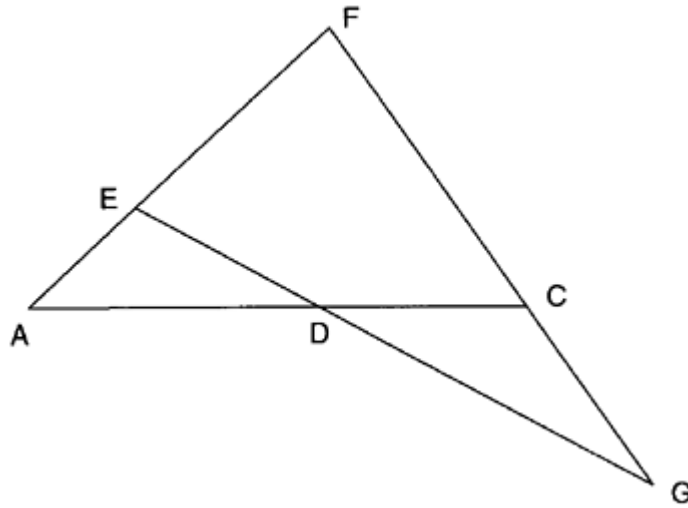
Figure H

Here, to conclude this topic, is an illustration of the use of pictures as a logical device from Hilbert's *Foundations of Geometry.* Certainly one use of pictures in geometry texts is to help you see that something is true of the figures drawn. But another use, which is not noticed, which is indeed the use Hilbert puts them to in his *Foundations,* is to provide a way for the understanding to stand in for logic, albeit not here in a way that would enable one to found a formal-logical law. Hilbert's proof of the theorem really consists of a demonstration of three lemmas by which it is clear from the picture that the theorem follows. The lemmas give all the facts that are needed from geometry. What is missing are the complex logical operations by which they would be formally brought together and made to yield the theorem, operations that involve, among other things, a lot of fussing around with quantifiers. By virtue of our understanding of the content of the propositions of the theorem and the lemmas, the picture enables us to see how/why the lemmas yield the theorem and, indeed, furthermore, how and why the axioms yield the theorem.

Theorem 3. For two points A and C there always exists at least one point D on the line AC that lies between A and C.

Proof: By virtue of the axiom which says that there exists three points not on a line, there exists a point E not on the line AC.

By virtue of the axiom which says that for any two points A, E, there is a third point F on the line AE, E between A and F, there exists on AE a point F such that E is a point of the segment AF.

By the same axiom and the axiom that says of any three points on a line there exists no more than one that lies between the other two, there exists on FC a point G, and G does not lie on the segment FC.

Given the axiom which says, if three points A, B, C do not lie on a line and if *a* is a line which passes through a point of the segment AB, the line *a* also passes through either a point of the segment AC or a point of the segment BC, it follows that the line EG must then intersect the segment AC at a point D.

The proof consists of actually four lemmas. These lemmas are put together via Figure H to yield the construction of the point required by the theorem. Here the picture stands in for that complex and messy formal-logical construction that would bring the four lemmas together to yield the theorem. The understanding that yields the aptness of the picture stands in for the formal logical construction. In any case, these are some of the ways in which the understanding stands in for the formal logic. In fact, so compelling and in its own way rigorous is Hilbert's proof that we might very well have grounds to suspect something wrong with any formallogical analysis which led to the claim that the theorem did not logically follow from the axioms.

II.
WHY POSSIBLE PROOFS ARE ACTUAL PROOFS AND THE NECESSITY OF MATHEMATICAL TRUTHS

It could be said that mathematics is the activity of looking for problems which could be completely solved in one's head, of trying to solve and actually solving such problems more or less or in principle in one's head, and, most splendidly, of throwing oneself at problems for which it could turn out that no one's head is quite big enough. It is a not altogether unwarranted suggestion that mathematics had its origins in the discovery that some practical problems which could have been solved by

observation could also be solved, somehow even more powerfully or cogently or compellingly, just by thinking, by puzzling things out. For example,

Of any three (normal) people, at least two will be of the same sex. This conclusion could have been reached by some sort of inductive generalization after conducting an empirical study of a large number of groups of three people. Or it could have been reached by a moment's *a priori* reasoning (e.g. "Given three people, in the worst possible case one person of the three would be male, one other female, but then the third would be either male or female, in either case giving two males or two females …").

Perhaps, indeed, mathematics had its origins in a recreation or an agon, in contests in which participants contrived and solved those riddles without pun or metaphor which we call "brain teasers," in puzzles that could be solved just by puzzling them out. If mathematics most broadly conceived is indeed the discipline of contriving and solving, ever more elaborately and systematically, problems that can be solved by wit alone, by just puzzling out without having to go out into the world and taking a look, then we have here ready-to-hand the perhaps most genuine sense in which "mathematical knowledge" is *a priori.* What remains is to attempt to characterize that sense in a philosophically careful way, and it is this that shall be attempted here. It is exactly because mathematical problems are problems that can be solved in one's head, and *for which the true or right solutions can be seen to be the true or right solutions entirely by wit alone through that device called "proof,"* that mathematical knowledge is *a priori* knowledge and is strongly protected against those of an empiricist *cum* naturalistic bent who would seek to deny this.

It has turned out that in order to articulate, develop, that just now roughly characterized sense in which mathematical knowledge is *a priori,* it is useful to introduce the concept of a certain sort of ideal language, that of "an *a priori* grounded language." Naturally occurring mathematical languages tend to fall short of this ideal, while certain artificial languages, in particular languages of constructive mathematics, seem to match the ideal.

A priori groundedness

"Warrant" as used here is always a warrant for the assertion of a sentence. The existence of a warrant warrants, grounds, authorizes, sanctions, certifies, validates…the asserting of the correlated sentence. Generally, a warrant for asserting a sentence does not assure the truth of the sentence but is sufficient to give an air of reasonableness to its assertion.

A language L as a whole has the trait of being *a priori* grounded if, and only if, there is no possibility of there being a warrant for the negative[5] Negs of a sentence s of L if it is possible that there be a warrant for *s.* If s could have a warrant, then Negs could not have a warrant. I do not mean only that its negative could not have a warrant at the same time, but that it could not (possibly) have a warrant at all. *Nothing could* warrant the assertion of Negs—*nothing* could make it reasonable to assert Negs.

Count as necessary and sufficient conditions for there being a possible warrant (or, better, that there possibly is a warrant) for the assertion of a sentence *s*

> 1.1 that we could in principle have a description so finely detailed that we could see from the description that,
>
> 1.2 were there something fitting that description, it would serve as a warrant for the assertion of the sentence *s,* and
>
> 1.3 the description misses out on no trait or aspect or attribute of the warrant by virtue of which it is a warrant for asserting that *s*.

Letting *"s,"* as noted, range over sentences of a language L,   over descriptions (not usually in *L*), write W[, *s*] for the relation of *warrants.* Then (keeping in mind that saying there is a   is to say that there is a description and not to say that there is something that fits the description),

> 1.1 if W[, *s*], then there is a possible warrant for asserting that s (although there need not be an actual warrant).
>
> 1.2 if W[, *s*] and there is something *x* that fits the description , then *x* is a warrant for asserting that s.
>
> 1.3 [*a priori* groundedness] L is *a priori* grounded if, and only if, for all sentences s of *L,* if there is a   such that W[, s], then there is no description   such that W[, Negs].
>
> 1.4 [*a posteriori* groundedness] L is *a posteriori* grounded if, and only if, for some sentence *s* of *L,* there is a   such that W[, *s*], and there is a   such that W[, Negs].
>
> 1.5 a sentence s of *L* is *a priori* grounded just in case for some , W[, s], and there is no   such that W[, Negs].
>
> 1.6 a sentence s of *L* is *a posteriori* grounded just in case for some , W[, *s*] and there is a   such that W[, Negs].

The idea that there ought to be such descriptions available if a warrant is to be possible is a common and natural one; for example, Phillip Kitcher uses the presumed unavailability of such descriptions as part of his arguments against intuition as a source of *a priori* knowledge:

If someone proposes that intuition be divorced from the sensuous then we have a right to ask for a description of the process of intuition which will enable us to identify it and to determine whether it can serve as an a priori warrant for

mathematical beliefs. Lacking a description of this kind, it is unclear that the proposal amounts to a theory of mathematical knowledge at all, much less a theory of a priori mathematical knowledge.

<div align="right">(Kitcher 1984:52)</div>

There are certainly complications arising with this notion of the possible warrants, the idea of the   , the nature of the relation (s?) W, and so on, which would have to be treated on a fuller account. There is, for one, the just encountered problem of whether it is a reasonable demand that warrants be elaborately describable. Another problem is that of how descriptions   can come to be found to bear the relation W to sentences *s.* Is it that we derive such descriptions on the basis of our understanding of *s,* for example, unpacking verification conditions out of that understanding and then imagining different ways in which they might be filled? Ought general methodological principles of rational inquiry, such as cultivated by, for example, Isaac Levi (1973) play a systematic role in deciding for what pairs of warrant descriptions and sentences the relation W holds? And ought there not to be more said about what sorts of descriptions could compose the   s? Surely not, for example, descriptions "the circumstance that *s* is true." Then there could be no *a priori* grounded languages since we would have "the circumstance that *s* is true" as a possible warrant for *s* and "the circumstance that Negs is true" as a possible warrant for Negs. It is easy to banish such cases by various devices, but difficult to work out a criterion that would allow in descriptions only of the sort that from them you could see that such a thing that would fit the descriptions is possible (the description would have to be something like a full imagining of the thing). But, fortunately, for the applications I will make of these notions here, I will not have to go much into these problems.

In any case, the notion of an *a priori* grounded language is clearly a strong notion. With such a language L we can associate a theory T_L consisting of all those sentences of L having a possible warrant. T_L would be a theory which could have no competition since any actually grounded theory would be a subtheory of T_L. In this case it would seem that we could just go ahead and assert T_L. It would seem—although this must be discussed below—that where there are no alternative or disjunctive possibilities (as there are not when, for example, there are no possible grounds for Negs), no possible way of finding for any sentence *s* of T_L that the assertion of Negs could be reasonable, the discovered possibility will suffice. In this case we may then speak of T_L as an *a priori* grounded theory. That is, in an *a priori* grounded language L there is a warranted theory T_L such that

1.1 any other warranted theories in L are subtheories of *T1;*

1.2 any theory in L which is not a subtheory[6] of T_L could *not possibly* be warranted;

1.3 the theory T_L is warranted by default. (There is not even a possibility of there being grounds for the assertion of any sentences of L not in *TL.*)

The question of pressing interest is, of course, whether or not mathematical languages are *a priori* grounded. If they are, then we would have a way of understanding the sense in which mathematical knowledge is *a priori* knowledge, namely, the mere description (which may be empty)   such that W[, *s*] suffices for asserting that *s.* We would also have a sense in which mathematical truths are necessary on some assumption that truth is assimilated to proof,[7] since if there is a possible proof for *s,* there would be no possible proof for Negs (it is not possible that *s* is not true).

Consider that *a posteriori* groundedness of sentences seems to be characteristic of empirical propositions. We must look to experience to decide among them because we are presented with exclusive alternatives. For a sentence s vulnerable to empirical grounds for its assertion, we can usually think of what would confute *s* and of what would refute *s;* and we can think out, characterize, describe, those possible grounds or warrants *in great detail.* Thus one might describe in thoroughgoing detail how a (perhaps nonexisting) meter is constructed, connected, and calibrated so that one can see from this that, for some sentence *s,*

(c) if the meter registers significantly below 2.72, then *s,* but

(d) if it reads significantly above 2.72, then Negs.

In the case of the description of the meter etc. together with the antecedent of (c) we have a description   such that W[, *s*] whereas in the case of the description of the meter etc. together with the antecedent of (d) we have a description   such that W[, Negs].

The circumstance of mathematical propositions differs quite markedly from that of empirical propositions. Warrants in mathematics usually come in the guise of proofs. Thus a typical possible warrant is given by a thoroughgoing description   of what? Of "something" such that, were it actual, it would be a proof (see 1.1– 1.3 above). But in the case of mathematics, having in hand such a thoroughly detailed description of a "possible" proof would be tantamount to actually having in hand a proof! Here is a description of a possible warrant (a possible proof) for the assertion that the *n*th root of 2 (*n*>1) is irrational:

It is said that the natural numbers are the numbers 1, 2, 3, 4, …, and the (positive) rational numbers are given by expressions of the form

$$\frac{p}{q}$$

where p and q are natural numbers with no common factors.

It is pointed out that every natural number can be uniquely written (up to order) as the product of natural numbered powers of prime numbers, so that

$$\frac{p}{q}$$

can be taken to be in the form

$$\frac{p_1^{n_1} \cdot \ldots \cdot p_k^{n_k}}{q_1^{m_1} \cdot \ldots \cdot q_j^{m_j}}.$$

It is then observed that

$$\left[\frac{p_1^{n_1} \cdot \ldots \cdot p_k^{n_k}}{q_1^{m_1} \cdot \ldots \cdot q_j^{m_j}} \right]^n = \frac{p_1^{n \cdot n_1} \cdot \ldots \cdot p_k^{n \cdot n_k}}{q_1^{n \cdot m_1} \cdot \ldots \cdot q_j^{n \cdot m_j}}$$

and then noted that the numerator and denominator will still have no common factors. From which it is to be noticed that if

$$\frac{p}{q}$$

is a rational number which is not a natural number (q 1), then raising it to the nth power cannot yield a natural number. It will then be noticed that it follows immediately that if a natural number does not have an nth root among the natural numbers, it does not have an nth root among the rational numbers.

It will then be noticed that 2 can have no nth root among the rational numbers for it has none such among the natural numbers, so that, as now can be said, the nth root of 2 is irrational.

Now here is the important point. To understand this description sufficiently to answer the question,

Could this possibly be a proof of the considered proposition? That is, for this description as , does it hold that
W[, *The nth root of 2(n> 1) is irrational*]!

is in effect to have seen a proof of the proposition. We would therefore certainly not expect to have such a detailed description of both a possible proof and a possible disproof of a theorem! It is as if a proof need only have a formal cause, needing no material cause.

We would then certainly expect mathematical languages to be *a priori* grounded. In particular, if one of our living mathematical languages were discovered to be *a posteriori* grounded, this would be indicative of a flaw, for then we could in effect prove of some sentence of that language both itself and its negative (granting that, as lightly argued for above, to have a description of a proof is tantamount to having a proof). Roughly, the reason that mathematical languages should have this trait at all is that mathematics is spawned by cultivating 'the field of problems that can be solved in one's head, by wits alone. These turn out to be problems for which the solution can be seen to be the solution by virtue of understanding alone. Proofs are just artifacts of the understanding. This is why possible proofs can stand in for actual proofs, so that the difference between actual and possible proofs is the difference of whether or not quotation marks are in place. But for a possible proof, exactly because proofs are an artifact of the understanding, to understand what is between the quotation marks sufficiently to see that it would be a possible proof of the proposition at issue is just to have sufficient understanding to find that, really, we must say that the proposition is proved.

This said, a much more delicate treatment of "possible proof" is needed. For the remainder of this section I will set out some ideas toward such a treatment, and the next section (III) will continue the matters a bit further.

Relative *a priori* groundedness and proofs

To have a warrant for asserting something is not generally to thereby have completely adequate assurance of truth—having good grounds for asserting something does not preclude the possibility that what is asserted is false. The considerations above were based on the identification of the possible warrants of assertion of sentences of mathematical languages with possible proofs. It is a very interesting, distinctive, and little noticed trait of proofs that a possible proof (as we have characterized

possible proofs in terms of descriptive language) is tantamount to an actual proof, and this has some interesting consequences for understanding the sense in which mathematical knowledge is *a priori* knowledge—it is *a priori* because possibility does suffice. At the same time the notion of an *a priori* grounded language gives us an idea of the sense in which mathematical propositions are necessary—if there being a possible proof for s entails there being no possible disproof of *s;* something that is possibly true (under these circumstances) is then not possibly false.

However, our identification of possible warrants with possible proofs is complexly problematic (although sorting things out can be rewarding). Warrants as understood generally do not necessarily clinch truth (otherwise we should have, for example, few or no empirical warrants). Now proofs are at least supposed to clinch the truth of what is proven. So might we not have warrants for mathematical propositions that are not proofs or, for that matter, proofs that do not clinch truth? About the latter, whatever the attitude toward proofs in mathematical practice (the ordinary mathematician supposes them to clinch truth), in fact proofs cannot be said with certainty to clinch truth in the sense of being absolutely beyond criticism, absolutely airtight. For example, our proof that the *n*th root of 2 (*n>*1) is irrational certainly has many gaps, including the there unproven assumption of the unique (up to order) representations of natural numbers by products of powers of primes. It is so familiar a point as to be a banality of the subject that, no matter how madly we try to fill in the gaps, there will be something unproven (but see Section III below). On this usual line of thinking, then, a rigorous proof will contain unproven assertions, which assertions are presumed to be based on warrants other than proof (typically presumed warrants—intuition such as "self-evidence" or considerations of fruitfulness of assumption). Our conviction that mathematical languages, if not desperately flawed, are *a priori* grounded depended on our identification of possible warrants with possible proofs, but these last considerations suggest that there are possible warrants occurring in mathematical practice that are not proofs. Indeed, John Stuart Mill, Kurt Gödel, Hilary Putnam, and many others have allowed that there are possible warrants that are not possible proofs. Gödel, for example, was open to intuitions providing warrants and to the satisfaction of an albeit not precisely formulated criterion of highly fruitful consequences as providing warrants for new axioms. Putnam, drawing on the work of Polya, has pointed out that inductive evidence is sometimes available for mathematical propositions.

It takes no heavy thinking to see that, indeed, not only can we have warrants that are not proofs, but we can have possible warrants (which are not possible proofs) for some sentences *s* of mathematical languages and also such possible warrants for the Negs, so that after all they are not *a priori* grounded but rather *a posteriori* grounded (see 1.4 , 1.6). We might, however, introduce the idea of a *relatively a priori* grounded language. Although a language *L* may be such that it is *a posteriori* grounded, it may be feasible to impose standards on warrants so that, in fact, for any sentence s only the possible warrants of it or the possible warrants of Negs, but not both, fall within the standards. This is perhaps a rather "ho hum" idea since it might be thought that any language where warranted assertion is an issue might be reducible to a relatively *a priori* grounded language by the imposition of some standards or other. To make the idea interesting we ought to select *peculiarly mathematical standards.* And indeed there is a natural choice, namely, possible proof! *The problem we immediately confront is to distinguish proofs from other sorts of warrants, to distinguish possible warrants which are possible proofs from other sorts of possible warrants.* But no sooner do we confront this problem than a serious step toward a solution to it suggests itself.

Possible proofs are those possible warrants given by a   for which   suffices for actuality, having, as we shall say, *the PAproperty* (possibility is sufficient for actuality property): to have a description   which is a possible proof of *s* and to understand it sufficiently to see that W[, *s*] is in effect to actually have a proof of *s.*

An *ideal mathematical language* is such that

1 if   is a possible warrant for a sentence *s* (of that language) and   has the PA-property, then there is no possible warrant with the PA-property for Neg*s*, and

2 for every sentence *s* or Negs, there is a possible warrant   with the PA-property.

The mathematical languages in which we operate at least have the trait that we cannot decisively show that both 1 and 2 hold (one way of demonstrating 1 holding would be to show the consistency of T_L for the language *L* of set theory—see 1.1 −1.3 above); but, of course, it may in some instances be the case that one or the other do not hold. Roughly speaking, the language of set theory is a case in point. Of the continuum hypothesis *CH* we do not know whether or not there is a possible proof of *CH,* a possible proof of Neg*CH,* or a possible proof of *CH & Neg CH.*

There must be some further qualifications on what counts as a description with the PA-property. For example, the description,

One visually sees a figure shaped like   .

certainly has the PA-property relative to possible warrants for the sentence,

There is a figure shaped like a triangle.

We can rule out such descriptions by the demand that the warrantability of the assertion of a sentence not be dependent on the existence of a description of a possible warrant for it (although we would still require for possible warrants that a description of the required sort—see 1.1–1.3 above—be available).

All this said and done, there might well by now be grave doubts about actual mathematical proofs having the trait of lending themselves to descriptions having the PA-property. After all, did I not point out that it is even a banality of the subject that proofs have holes, that those holes are at best filled by appeal to warrants that are not possibly proofs, such as the appeal to intuition? Consider, indeed, the last. Descriptions giving us at least possible mathematical warrants in which there is an appeal to intuition by way of warranting some assumptions, from which perhaps elaborate deductive inferences are made, will have as a clause in the description something to the effect "Here, to begin with, find or see that *s*" or "There is an actual intuition by virtue of which it is seen that *s*." These mere descriptions, however, will hardly do for a warrant. Possibility is not as good as actuality. One needs to actually have the intuition, to actually find or see that *s*.

But, I challenge, might such a circumstance indicate that in going on to include such a clause as "There is an actual intuition by virtue of which it is seen that *s*" in the description supposedly constituting a possible warrant, a possible ground, *one has gone beyond proof?* I think that there is a way of making *yes* to this question the proper and even natural answer. Isaac Levi (1973:2– 6) (among others) has distinguished between *local* and *global* justification, and I would propose that, in mathematics, proofs are the central tool of local justification:

> In science, justification of belief is demanded only when the need for such justification arises in the context of specific inquiries. Philosophers have often been discontent with justification in this "local" sense. Like Descartes or some of the contemporary writers who worry about choosing between "conceptual schemes," many philosophers occupy themselves with efforts at the more "global" justification of the totality of beliefs held at a given time.
>
> (Levi 1973:1)

> The evident (if such there be) is evidence in local inquiry because all parties to the inquiry agree that such beliefs do not stand in need of justification.… To include H in the evidence requires that (a) at the time of the inquiry H be believed true, and (b) at that time, critical scrutiny of H in the light of new evidence be considered pointless.…
>
> (Levi 1973:5)

However, it will be part of my considerations that mathematics is peculiarly distinct among modes of rational inquiry. Levi had gone on to remark,

> the central problem of a theory of local justification or rational belief is the establishment of criteria for determining which of the relevant answers to a given question, on the evidence available, is the best. These criteria do not determine what questions are raised or what evidence is available. Rather they reveal features of legitimate inference that are invariant over broad categories of local inquiries, regardless of the questions and the evidence peculiar to each individual problem.
>
> (Levi 1973:5)

What is peculiar about mathematics is that it is a tacit, always operative, ideal of mathematical thinking to be free of the need for such a theory of local justification or rational belief—the tacit ideal is that if, for a piece of ongoing, active mathematics, there is a genuine use for a theory of local justification good for scientific or rational inquiry generally, then that piece of mathematics is deeply flawed, most likely in its very conception, in its very language. In particular, I would want to suggest that, whereas this would not be an especially reasonable ideal outside of mathematics, it is a reasonable ideal within mathematics. One way of expressing that ideal is to say that the language of an ideal piece of mathematics ought to be relatively *a priori* grounded, where the standard of admissible warrants is that they have the PA-property.

Now, do actual proofs have the PA-property?

Answering this question in any thorough, rigorous way will require rather extended phenomenological studies, at the very least in order to learn how to formulate the PA-property more exactly. In the remainder of Section II and in Section III, I want to initiate such studies in the simplest cases.

We learn the plus and times tables for the nonnegative integers 0, 1, …, 9 and then use standard algorithms for summing and multiplying larger numbers via arabic, 10-ary numerals, algorithms reducing such summing and multiplying to those plus and times tables. Those of us who were not good students, had bad memories or were otherwise disrespectful of authority had to work out or keep working out the sums and times tables for ourselves. The sums we worked out using fingers or desks in the classroom or panes of glass in the window or trees outside the window. The times tables were a bit trickier, especially if one did not have a sense of the commutativity of multiplication, for example, working out 7 times 8 (having understood 7

1 2 3   1 2 3 4 5 6
1 1 1 + 1 1 1 1 1 1 =

1 1 1 1 1 1 1 1 1

1 2   3 4 5 6 7 23   27

1 2 3   1 2 3 4 5 6
1 1 1 + 1 1 1 1 1 1 =

1 1 1 1 1 1 1 1 1

1 2 3 4 5 6 7 8   9
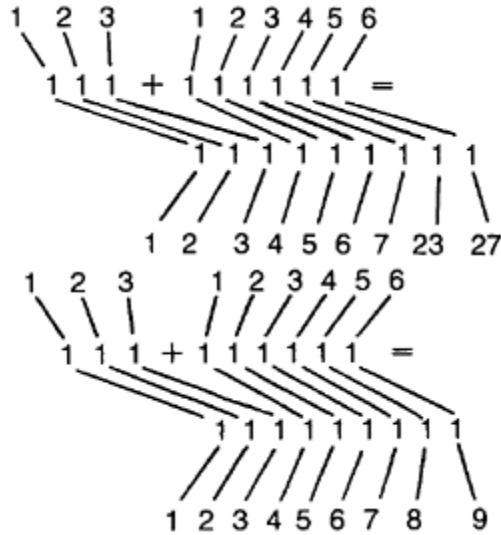
times 8 as eight 7s) by having noticed that 10 times 8 (eight 10s) is 80 and then subtracting eight 3s. In any case, for summing, computing sums depended on correlating 1-ary numerals and 10-ary numerals. It is safe to say that, down to the dawn of time, sums, whatever the numeral system, were computed in terms of things as 1-ary numerals counted in the numeral system of interest (10-ary arabic for us). The elementary arithmetic truths that were thus given are of such a nature that, as Hilbert in effect observed, a more rigorous proof (such as, for example, Frege was after) would have to rely on principles or a proof procedure which was more problematic, that is, it would not be so much a test of the elementary arithmetic truths (composing the sum tables, say) that they followed from such principles and proof procedures, but, rather, it would be more a test of those principles and proof procedures that they yielded up the arithmetic truths (and not at all their contradictories).

Consider this supposedly possible proof that 3+6=27:

Description of the proof that 3+6=27:
The numerals from one to twenty-seven are: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27.

Now consider this description of the supposedly possible proof that 3+6=9:

Description of the proof that 3+6=9:
The numerals from one to twenty-seven are: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27.

Thinking through what *is described* in the *first case* it becomes brutally clear that what *is* described could not be a warrant for 3+6=27 since 1, 2, 3, 4, 5, 6, 7, 23, 27 is not an initial sequence of the 10-ary arabic numerals and *so does* not rightly count the components of the 1-ary numeral 111111111. Therefore after all the description *does* not amount to a possible proof; the description does not give us a possible warrant. But in the case of the second description, one not only finds that such a thing would warrant asserting that 3+6=9, but, after checking it over a couple of times, one cannot imagine a way in which it might be flawed, and here one *must mean genuinely or fully imagine* so that one could actually test out whether or not the flaw is present. The first description gives us a sense of the sort of thing that could go wrong (miscounting or abuse of numeration), but nothing like that has gone wrong in the case of the second. Most importantly, one cannot not imagine a possible warrant for Neg*3+6=9,* such as a possible warrant for 3+6=27, where some miscounting has not occurred—which is to say that one cannot imagine a possible warrant for Neg*3+6=9.*

In the case of the second description, it clearly has the PAproperty.

III.
PROOFS WITHOUT UNPROVEN ASSUMPTIONS

The thesis here is that by and large proofs in mathematics do not involve unproven assumptions. Some proofs contains assumptions, but it is also the case that they are typically proved elsewhere. This thesis will be rather less successfully articulated and developed than the topics in Sections I and II. The topics of the first two sections could have been developed—

and I hope that this is clear— in considerably more detail and over a far wider range of example and consideration. The insights and strategies of analysis important to the development of Section III are not in place—there is rather more phenomenological work yet to be done here. The major block is that philosophers have not yet secured a sufficiently powerful and subtle sense of the understanding. Mainly I will consider some very simple (but nevertheless not easy) examples, hoping that they might prove suggestive to the reader. But let me be just a bit clearer about what the thesis is.

   Given a mathematical proof. Typically this proof will depend on theorems or lemmas that have been proven elsewhere. The thesis being considered here in Section III is that, typically, if we trace backwards we will find that, in the end, there have been no unproven assumptions. I actually mean by this something which might look weaker, but, if you think about it, is not— *there is nothing in the proof that needs to be proved.* Of course what I have in mind is that there are no postulates underlying mathematical proof. Mathematical demonstration is, typically, not founded on unproven assumptions. Of course, it is sometimes of interest to a mathematician to derive consequences of, say, the postulates of mathematical ring or field theory (or whatever). So one can consider, perhaps, that here, at least, the mathematician is proving something on the basis of postulates. But is this so? It might be thought that postulates framing natural numbers or sets or calculable functions might somehow be evident on the basis of certain semi-mental entities usually called "concepts." And this is not implausible. But the concepts of mathematical groups, rings, fields, etc., emerge from the isolation of significant mathematical structures and really cannot be thought of as somehow available in advance through concepts giving a basis on which axioms characterizing those structures would be evident. But what is important from our present point of view is that the proof that such and such follows from, say, the field axioms is not the same as a formal derivation from the axioms (gapless formal-logical derivations are extremely rare and usually far too long) but, rather, *a proof that either there is a formal derivation or else a proof that the theorem follows from the axioms,* as in Hilbert's *Foundations of Geometry.* Such proofs will typically not contain unproven assumptions— assumptions in need of proof, or so I claim. Even in the case that what is at issue is a formal-logical system, a given formal-logical derivation will not count as a proof of its own existence; rather, that it is a correct formal-logical derivation is something that must be demonstrated, and certainly not by a formal-logical derivation! By what then? by a series of observations that lets us see/find that it is a formal-logical derivation of a sentence from the axioms according to given rules of inference. Where there is a formal axiomatic system, the genuinely mathematical theorems and proofs are not formal derivations from the axioms (generally, mathematicians do not want to see such things!) but rather proofs that there exists such and such a formal derivation of such and such from such and such axioms according to such and such rules. Far from formal derivations being adequate to represent proofs, they themselves as it were need proofs—need to be proved to exist![8] For such formal derivations, such proofs may differ more or less only in complexity from proofs of elementary arithmetical identities. For this reason and also just because it seems we must begin by looking at such simple cases, I want to continue the studies begun above of proofs of elementary arithmetic identities, claiming the superiority of finger-counting demonstrations.

   About such demonstrations it will seem that there is nothing there in need of being proved.

   The treatment of demonstrations of simple arithmetic identities to be sketched really requires something like a philosophy of arithmetic behind it if it is to have much force, so some remarks in this direction might help. In his *Foundations of Arithmetic,* Frege had argued that numbers are substance-like entities, "self-standing" entities (Section IV: "Every single number is a self-standing object."). Dedekind (1963:68) had argued that numbers are the abstract relata of a certain relational structure—there is nothing more to them than that they are such relata. Neither would have thought it at all correct that numerals are numbers. But the last is the view of Lebesgue (1966:12–19), who was nevertheless fully and acutely aware of those philosophical views that made this identification seem absurd.

> We can easily imagine…the need to compare two collections led man to *counting,* that is, the comparison of two collections with a single standard collection—the collection of the words in a certain phrase. These words are called *numbers.*

In our Hindu-Arabic 10-ary numerals,

$$0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10\ 11\ 12\ldots$$

each numeral encodes a standard collection, *a number* of things, a standard number of things. For example 7 encodes the collection,

   {0 1 2 3 4 5 6}

and 0 encodes the collection

   { }.

A number is an answer to the question,

   How many?

6 + 5

1 2 3 4 5 6   1 2 3 4 5

+   =

1 2 3 4 5 6 7 8 9 10 11

QED   11
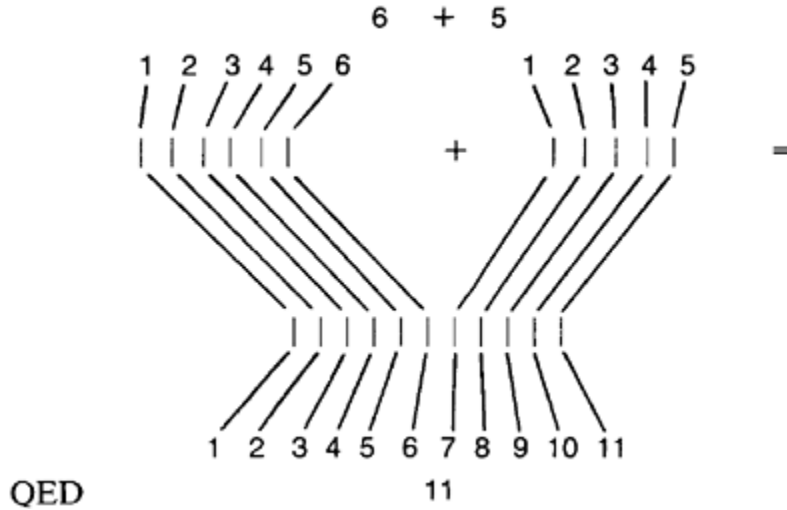
The as it were most basic answer to this question is,

This many: (and now give a collection).

How many hands do you have?

This many: {0 1}

or

2 (for short)

Lebesgue remarks,

> It is not the word number that we need to explain but rather the sentences in which that word appears. For instance, consider the following sentences: Two collections have the same number; two collections do not have the same number. Now, this is precisely what was explained at the outset in describing the operation of numbering a collection. Thus, we remove any reason for metaphysical fears.
>
> (Lebesgue 1966:16)

In the spirit of Lebesgue's conception of number, here is a proof of 6+5=11 having no assumptions in need of proof:

Beyond exploring such examples in greater depth one might turn to Hardy and Wright's *An Introduction to the Theory of Numbers.* There, number theory is developed without the formulation of axioms. But do they covertly appeal to axioms? Just as Hilbert did not covertly appeal to rules of inference, but, rather, his inferences rested on an understanding, an insightful understanding, which was prior to, independent of, rules of inference, but on which rules of inference might be based—intuitive logical thinking he called it—so Hardy and Wright do not covertly appeal to axioms, but, rather, their proofs rest on an understanding which is prior to axioms. But just as the patterns of deductions given by Hilbert are not structured as if by rules of inference (we have seen how Hilbert's quite certain use of pictures results in deductions significantly evading such patterns), so there are no neat axioms in Hardy and Wright standing just off-stage and cueing the proofs. There is, rather, a fluid understanding of number variously deployed. A close look at the opening of the text gives some idea of the way in which understanding stands in for axioms. Well, just consider a part of the proof of

Theorem 1. Every positive integer, except 1, is a product of primes.

Proof: Either n is prime, when there is nothing to prove, or n has divisors between 1 and n. If m is the least of these divisors, m is prime; for otherwise…

Indeed, throughout the work Hardy and Wright speak of the least positive or nonnegative integer having some property. They are *not,* however, appealing to "the least number principle," even tacitly:

[LNP] Any property holding of a positive integer holds of a least positive integer.

or

Any non-empty set of positive integers contains a least integer.

It might be thought that they are appealing to the understanding of the nonnegative integers being totally ordered, and to something more understood than formulable in a way adequate to that understanding. The curious thing is that, if we try to rigorously formulate this idea, what is supposedly understood, we have, it seems, to draw on a concept that is not first-order expressible, namely *Every nonnegative integer is preceded by a finite number of integers.* However, the notion of a finite number or collection or set at issue here is such that *it results* from an attempt to capture something otherwise understood, the peculiar trait of the collections encoded by the (say) Hindu-Arabic decimal *numerals* for the nonnegative integers, a trait instantly or immediately handed over by the nature of the *numerals* and that makes it available for Hardy and Wright to say, without so much as hinting at the presence of a derivation from an unproven postulate for which "If *m* is the least of these integers…" would be elliptic. It is an obvious trait of the familiar Hindu-Arabic 10-ary numerals that for any numeral there is a next largest and for any numeral except 0 there is a next smallest, and that going from one numeral stepwise to the next largest one can get to any larger numeral, and that going stepwise to the next smallest one can get to any smaller numeral. This makes it obvious that there is a least number qua numeral having any property had by some number qua numeral. Now this understanding is logically flawed, as Frege rather elaborately pointed out. One should not mistake numerals for numbers, or confuse the two. At the same time, in doing arithmetic or number theory, it is not at all likely that this ostensible confusion leads to any arithmetically false theorems. Furthermore, the attempt to replace an understanding based on numerals, for example, replacing such notions as "going stepwise to the next smallest one can get to…," by a more objective wholly mathematical notion (e.g. in terms of order, logically characterized finiteness, etc.) must answer to that logically more confused understanding based on obvious traits of numerals, or the taking of numbers as numerals. That logically confused understanding is nevertheless cogent enough to generate number theory. It is only when we replace that understanding with logically, objectively formulated principles, such as the least number principle or mathematical induction, that we end up with assumptions in need of proof, with a need for axioms or postulates themselves unproven and, as in the case of arithmetic, if evident at all, most likely evident (which is not to say objectively certain) by virtue of reference back to a more confused, but cogently managed, understanding. This is not the same thing as saying that that understanding is above criticism, and there is no assurance that such understanding may fail to keep the mathematics based on it from falling into real mathematical confusions as Euler's deployment of infinitesimals eventually led to irremovable mathematical confusions about higher-order differentials.[9] But it is to point out the sort of basis all but ubiquitously present in mathematics enabling proofs to go through without unproven assumptions.

While it is even obvious, if one thinks about it for a moment, that such passages as cited in Hardy and Wright indicate a background of understanding which does not at all have the form of the presence of presumed derivations from hidden and unproven assumptions—an understanding making it possible for there to be proofs which contain no assumptions in need of proof—it will be a highly significant act of presumably phenomenological philosophy to discover how to satisfyingly characterize or represent that understanding (rather than just appealing to it), even a great philosophical advance to learn how to characterize it satisfyingly, for it seems that the only conventional way of representing it is to radically misrepresent it by just setting out postulates or axioms and derivations therefrom. It is worth noticing that the otherwise extremely acute Frege when confronted with just this characterizational problem lapsed into one of his few genuinely idiotic remarks, namely "everyone recognizes the same geometrical axioms, even if only by his behavior…" (Frege 1967:35). (It is the "even if" that is so troublesome.)

Summing up, it might be said that reasoning in arithmetic, centering, for example, on assertions "take the least number such that…," is justified on the basis of the understanding of the system of numerals; the very fact that the logical analysis of such assertions leads to second-order assertions shows that the logical account of the least number principle does not give an adequate representation of our understanding, our understanding on which arithmetic or number theoretic reasoning is based, and so must be understood as based on our more intuitive understanding of our numerals as sketched above. That is, logical representation is but one sort of representation and may not be adequate to represent the understanding on which our mathematical reasonings and proofs are in fact founded!

Finally, it is worth pointing out that in recent work W.W.Tait has been uncovering a way of construing proof in classical mathematics as construction, in this measure freeing proof from assumptions in need of proof (Tait 1986a, b). (However, it is questionable that this is a phenomenologically adequate account of proof rather than a rational reconstruction of it.)

NOTES

1 From the point of view of the present paper, perhaps the most useful work in English introducing some of the main ideas of Husserl's transcendental phenomenology in the spirit in which it was set out, a kind of baseline work to get one's thinking started, is Sokolowski (1974).

2 For the development of my particular way of practicing phenomenology (albeit much improved in the present studies) and other background, see Tragesser (1977, 1984).

3 Locke (1975), p. 672; see the amusing comments in Leibniz (1981), p. 477.

4 I must thank Hidé Ishiguro for pointing out that many philosophers, presumably because they fail to distinguish the formal from the informal in the senses set out above, virtually by reflex would assess Frege's argument or demonstration as using logic to justify logic and therefore not a sound argument/demonstration.

5 In order to avoid clutter (relying on the reader to see how things would go), I here and elsewhere avoid worrying about the circumstance where NegNeg*s* is not equivalent to *.s*.

6 I understand "subtheory" in such a way that any theory is a subtheory of itself.

7 See Tait (1986a).

8 An analogous point about the way in which intuition gets into more formalistic accounts of constructive mathematics is made by Charles Parsons.

9 See Bos (1974).

REFERENCES

Bos, H.J.M. (1974) "Differentials, higher-order differentials and the derivative in Leibnizian calculus," *Archive for History of Exact Sciences* 14:1–90.

Dcdekind, R. (1963) *Essays on the Theory of Numbers,* New York: Dover Publications.

Frege, G. (1967) *The Basic Laws of Arithmetic,* trans. and ed. M.Furth, Berkeley and Los Angeles, CA: University of California Press.

Hardy, G.H. and Wright, E.M. (1979) *An Introduction to the Theory of Numbers,* 5th edn, Oxford: Clarendon Press.

Harman, G. (1973) *Thought,* Princeton, NJ: Princeton University Press.

Hilbert, D. (1965) "Axiomatisches Denken," in D.Hilbert *Gesammelte Abhandlungen,* vol. 3, Bronx, NY: Chelsea Publishing, 146–57.

——(1971) *Foundations of Geometry,* 2nd English edn, trans. L. Unger (from the 10th German edn), LaSalle, IL: Open Court.

Hilbert, D. and Ackermann, W. (no date) *Principles of Mathematical Logic,* trans. L.M.Hammond, G.G.Leckie, and F.Steinhardt, with R.E.Luce, Bronx, NY: Chelsea Publishing.

Husserl, E. (1970) *Logical Investigations,* trans, from the 2nd German edn (1913) by J.N.Findlay, New York: Humanities Press.

Kitcher, P. (1984) *The Nature of Mathematical Knowledge,* Oxford: Oxford University Press.

Lebesgue, H. (1966) *Measure and Integral,* ed. K.O.May, London: Holden-Day.

Leibniz, G.W. (1981) *New Essays on Human Understanding,* trans. and ed. P.Remnant and J.Bennett, Cambridge: Cambridge University Press.

Levi, I. (1973) *Gambling with Truth: An Essay on Induction and the Aims of Science,* Cambridge, MA: MIT Press.

Lewis, D. (1986) *On the Plurality of Worlds,* Oxford: Basil Blackwell.

Locke, J. (1975) *An Essay Concerning Human Understanding,* ed. P.H. Nidditch, Oxford: Oxford University Press.

Parsons, C. (1986) "Intuition in constructive mathematics," in J. Butterfield (ed.) *Language, Mind and Logic,* Cambridge: Cambridge University Press, 211–29.

Sokolowski, R. (1974) *Husserlian Meditations,* Evanston, IL: Northwestern University Press.

Tait, W.W. (1986a) "Truth and proof: the platonism of mathematics," *Synthese* 69:341–70.

(1986b) "Against intuitionism: constructive mathematics in part of classical mathematics," *Journal of Philosophical Logic* 12:173–96.

Tragesser, R.S. (1977) *Phenomenology and Logic,* Ithaca, NY: Cornell University Press.

(1984) Hu*sserl and Realism in Logic and Mathematics, Ca*mbridge: Cambridge University Press.

ON AN ALLEGED REFUTATION OF HILBERT'S PROGRAM USING GÖDEL'S FIRST INCOMPLETENESS THEOREM

Michael Detlefsen

SUMMARY

In this paper it is argued that an instrumentalist notion of proof such as that articulated in Hilbert's notion of *ideal mathematics* is not obliged to satisfy the conservation condition that is generally presented as a constraint on his program. A weaker and more plausible soundness condition is then identified and shown not to be counter-exemplified by Gödel's First Theorem. In the final section, attention is focused on the question whether there is room within an instrumentalist framework for a noncumulative (in current jargon "nonmonotonic") conception of ideal proof or, equivalently, whether an instrumentalist theory might be better thought of as a *method for selecting* beliefs (theorems) than as a *fixed set* of beliefs (theorems). The potential importance of this issue for correctly judging the bearing of Gödel's Second Theorem on Hilbert's Program is considered.

I.
INTRODUCTION

For some time now, it has been a virtual constant of the literature on Hilbert's Program (HP) to maintain that Gödel's work demonstrates its untenability. The "demonstration" typically given is one which proceeds from Gödel's Second Incompleteness Theorem (G2) and the claim that HP requires the sort of consistency proofs that it (i.e. G2) rules out. However, more recently (cf. Kreisel 1976; Prawitz 1981; Simpson 1988; Smorynski 1977, 1985, 1988) it has become increasingly common to claim that Gödel's First Incompleteness Theorem (G1) affords a refutation of HP, and that this refutation is at least as good as (and perhaps even better than) that based on G2. Thus, one finds such claims as that "the First Incompleteness Theorem…effectively kills Hilbert's programme; the Second Incompleteness Theorem is merely a refinement" (cf. the introduction to Smorynski 1988), and "it was the *First* and not the Second Incompleteness Theorem that killed Hilbert's Programme" (cf. 1985:10).

Like the G2-based argument, the G1-based argument proceeds by producing a "requirement" for HP. This requirement is then shown to be ruled out by G1. In the case of the G1-based argument, however, the requirement is not one of consistency, but rather one stating that what HP identifies as "ideal" mathematics must be a conservative extension of what it identifies as "real" mathematics. G1 is then said to show that this conservation requirement cannot be met, the upshot being that G1 is sufficient, and hence G2 not necessary, for the destruction of HP.

We disagree with this argument. G1 does not "kill" HP, and thus does not diminish the critical importance that the G2-based argument has for any attempt to establish a link between the defeasibility of HP and Gödel's work. In saying this, however, we are not intending to suggest that the G2-based argument against HP is successful; indeed, we have argued at length elsewhere (cf. Detlefsen 1986) to the contrary. Our aim is rather to affirm the pivotal status of the G2-based argument and thus to focus attention on it. We are thus making a proposal concerning what we take to be the proper focus of work aimed at determining the effect of Gödel's work on HP. The objectives of the present paper, then, are first to show that the G1-based argument is mistaken, and second to indicate what the main issues are that must be addressed by any rationally convincing evaluation of the bearing of Gödel's work on HP.

In connection with the first objective, we focus our attention on the role which the G1-based argument assigns to the conservation condition, and on the other alleged constraints on ideal theorizing which are then supposed to lead to its violation. We argue that both the conservation condition itself, as well as the other constraints on ideal theorizing which are taken to underlie its violation, are mistaken.

Regarding the second objective, our concern is mainly with the different conditions which G1 and G2 place on the (encoded) notions of proof, provability, etc. The conditions required by G2 are not satisfied by various theories or "theory-like" arrangements of proofs and theorems which nonetheless *do* satisfy the conditions required by G1. Chief among these, so far as our interests are concerned, are various types of theories which incorporate consistency constraints into the very

conditions on proof, provability, etc. We call these theories "consistency-minded," and we maintain that some of them constitute plausible ways in which the Hilbertian might go about constructing his ideal theories.

There are two broad types of such systems: those introduced in Rosser (1936) (and studied and refined in Kreisel and Takeuti (1974), Guaspari and Solovay (1979), Visser (1989) and Arai (1990, on the one hand, and those introduced by Feferman (1960) (and studied in Jeroslow (1975) and Visser (1989)), on the other. Roughly speaking, the former (which we refer to as Rosser systems) begin with a class of would-be proofs and an -ordering defined on them, and require that, to be genuine, a would-be theorem must be determined neither to deny nor to be denied by any would-be theorem whose would-be proof precedes it in the given ordering of the would-be proofs. (Note that, since this requirement involves comparison of only the last line of the given would-be proof with that of its finitely many predecessors in the given -ordering, it is effectively executable.)

Similarly, the latter type of systems ("Feferman systems," in our terminology) makes use of a class of would-be axioms and an ordering defined on them, and demands that for a would-be axiom to be genuine it must be consistent with the whole class of wouldbe axioms which precedes it in the given ordering. This in turn gives rise to a notion of proof which requires that, to be genuine, a would-be proof must be such that its "largest" axiom (i.e. the one lying farthest out in the above-mentioned ordering) is consistent with the class of all "smaller" ones. (This means that, unlike Rosser systems, Feferman systems are not effective since, by Church's Theorem, there is no effective way of carrying out the consistency check they require. However, in Jeroslow (1975) (and again in Visser (1989)), ways of obtaining effective systems based on Feferman-like ideas are suggested. Jeroslow calls these systems "Experimental Logics" since their axioms are, in a sense, determined by a trial-and-error procedure.)

Because (at least some of) these consistency-minded theories seem to represent sensible strategies of theory construction for the Hilbertian, and because they at the same time violate certain of the conditions required for the proof of G2, they raise the question of whether these conditions are something to which the Hilbertian is committed by the nature of his enterprise. That these conditions hold for the usual sorts of systems but not for their consistencyminded counterparts should cause us to ask why the Hilbertian should be seen as committed to building his systems of beliefs in the usual static manner (where demonstrated consistency with previously accepted beliefs is not a condition on admitting a proposition as a new belief (i.e. theorem)) rather than in the more dynamic way suggested by the consistency-minded modes of construction. The answer, we believe, is that there is no reason; and this raises the question of whether G2 applies to Hilbert's Program *per se,* or only to those versions of it which needlessly restrict themselves to theory construction of the usual static variety.

We believe that this is an open question; and the primary objective of this paper is not to solve it, but rather to convince the reader of its seriousness, and to lay the philosophical groundwork necessary for a fruitful discussion of it. In this way, we hope to establish a general philosophical framework capable of serving as a guide to further work on the subject.

These, then, are the objectives of the paper. Its plan is as follows. In the next section we attend to preliminaries, setting forth a correct account of Hilbert's real/ideal distinction, which plays a central role in determining what sort of soundness condition it is appropriate for the Hilbertian to place on ideal mathematics. In Section III, we give an analysis and critique of the G1-based refutation of HP. Finally, in Section IV, we attempt to lend some perspective to the refutation of the G1-based argument given in Section III by showing what happens when attention is shifted back to the G2-based argument. Our belief is that the possibility of a consistency-minded conception of the Hilbertian's ideal theorizing may afford a way for the Hilbertian to carry out his program unimpeded by G2. At any rate, this is an alternative we think merits serious investigation and whose philosophical groundwork we intend to prepare.

II.
THE REAL/IDEAL DISTINCTION

In order to gain a proper understanding of the G1-based argument against HP, it is necessary to begin with a proper understanding of Hilbert's well-known (if not always well-understood) distinction between *real* and *ideal* mathematics. Hilbert himself seems to have intended this distinction to mirror the familiar distinction (of days gone by) between observation and theory in the natural sciences,[1] according to which observational evidence acts as a constraint on proper theorizing in this sense: no proper natural scientific theory can admit as consequences statements whose falsehood can be established by observational means. In this way, observation statements (i.e. statements whose truth is decidable by observational means) function to control theorizing in the natural sciences.

Construing the real/ideal distinction in parallel fashion, real proofs occupy the role of observational evidence, and ideal propositions the role of theoretical sentences, so that the real proofs function to control ideal theorizing in the same way that observational evidence controls theorizing in natural science. Hence, just as no adequate natural scientific theory is permitted to have observationally falsifiable consequences, so too no proper ideal theory is permitted to have consequences whose falsehood can be established by real (i.e. finitary) means.[2]

Hilbert seems to have viewed both natural scientific theory and ideal mathematics instrumentalistically. Thus, in drawing the analogy between ideal mathematics and theorizing in physics (1927:475) he says that

> a theory by its very nature is such that we do not need to fall back upon intuition or meaning in the midst of some argument. What the physicist demands precisely of a theory is that particular propositions be derived from laws of nature or hypotheses solely by inferences, hence on the basis of a pure formula game, without extraneous considerations being adduced.[3]

For Hilbert, then, ideal sentences do not have any genuine semantical or justificatory standing of their own. They do not express propositions but are rather part of a formal calculary device (the system of ideal proofs or derivations) whose purpose is the timely and efficient derivation of sentences (namely, real sentences) that *do* have such semantic and epistemic standing.[4]

Smorynski (1988) is a reaction to this way of understanding the real/ideal distinction. His idea is to replace "the old dichotomy between finitary propositions and transfinite, or ideal formulae" with a new "trichotomy" which distinguishes three types of propositions; namely, *real propositions, finitary general propositions,* and *ideal propositions* (cf. pp. 58–9). In this new trichotomy, the real propositions are identified with equations relating terms for primitive recursive functions and/or fixed numerals and their finite propositional combinations. Their character is described as that of being "directly contentual assertions verifiable by direct computation." Their epistemological significance is that of "affording a control on the results of formal mathematical proofs." Thus, their epistemological role is like that of experimental or observational controls in the natural sciences, and they therefore play the same basic role played by the reals in the old dichotomous scheme.

The elements of the third category of the "new" trichotomy— the so-called ideal propositions—are also conceived of in the usual way. They are said not to be genuine propositions at all, but rather mere formulas, and are said to bear the same relation to the reals as the theoretical elements of an instrumentalistically conceived natural science bear to the observation statements which pertain to it. Here, however, this relationship is apparently taken to be given not by the sort of *soundness condition* stated above (namely that requiring that the theoretical elements not generate any observationally falsifiable consequent), but rather by a *conservation condition* requiring that any observational proposition derivable by theoretical means also be provable by observational means. At any rate, Smorynski maintains that in order for a given ideal theory **T** to be adequate, all real sentences provable in **T** must also be provable by real means (cf. p. 64). Of course, since the reals on Smorynski's scheme are all finitarily decidable propositions, there is no essential difference between his condition of real-conservation and the usual condition of real-soundness.

In its first and third categories, then, Smorynski's trichotomy follows fairly closely the usual view of the real/ideal distinction. The novelty comes with his second category, the so-called *finitary general propositions* (or "fgps," for short). These are generalizations such as "for all numerals *n, n*+1=1+*n*." According to Smorynski, the fgps, like the real propositions, are supposed to be meaningful. Moreover, like the real propositions, they are said to be capable of finitary proof (albeit by what are called "schematic methods"). What distinguishes them from real propositions, says Smorynski (cf. pp. 59–60), is that they are *infinite conjunctions of* real propositions, and, as such, function as the "laws" of mathematical science. This is taken to be important because it is assumed that every genuine science must have laws (cf. Hilbert's (1925:376) comment that the science of mathematics is neither exhausted by nor reducible to equalities and inequalities).

Fgps thus seemingly have dual status; in their office as universal laws, they occupy the place of "theoretical" elements, while in their role as objects of finitary proof, they function as "controls" on ideal theorizing (and thus are to be included in the scope of the conservation condition as propositions which must be provable by finitary means whenever they are provable by ideal means).[5] We are thus left with the suggestion that the theoretical part of mathematics is a mixed bag, partly realist in character (owing to the genuineness of the finitary general propositions), and partly instrumentalist in character (owing to the merely "computational" status of the so-called ideal propositions).

Many questions concerning this trichotomous conception of the real/ideal distinction are left unanswered by Smorynski. (For example, if the fgps, like the reals, are subject to finitary proof, why is it only the latter and not the former which function as evidensory controls on ideal theorizing? If, on the other hand, the fgps do not function as controls, then why should ideal theorizing be required to be conservative with respect to them? Moreover, are the fgps propositions or prepositional schemata? And how, exactly, do they compare with the ideal propositions?) Of more immediate concern than this, however, are the inaccuracies and confusions upon which it is built.

The chief culprit in all this is the introduction of the "new" category of fgps, which Smorynski sees as being based on two grounds. The first is the aforementioned idea that every genuine science must have laws (cf. pp. 59–60). Smorynski takes Hilbert to have been referring to a realm of "laws" (namely, Smorynski's fgps) when he spoke of that part of mathematics that cannot be reduced to numerical equations. In another place, however (cf. Hilbert 1927:471), Hilbert put his point more carefully. There he characterized mathematics not as the adjunction of an equational part and a set of contentual laws (Smorynski's fgps) related to the equations as theory to data, but rather as the adjunction of real and *ideal* methods where the

latter are related to the former as theory to data. Indeed, he says explicitly that what is essential for a genuine science of mathematics is not a realm of contentual laws (like Smorynski's fgps), but rather the use of ideal methods ("scientific mathematics becomes possible only through the introduction of ideal propositions").

Thus it seems that what Hilbert had in mind when he spoke of a part of mathematics that is essential to its being a science and is not reducible to numerical equations is *ideal* mathematics, and not, as Smorynski suggests, a realm of contentual laws lying somewhere between ideal mathematics and elementary equalities and inequalities. Likewise, when he spoke of the equational part of mathematics as forming a control on the part of mathematics not reducible to it, this was just a loose, and somewhat sloppy, way of saying that the noncontentual part of mathematics (i.e. ideal mathematics) is to be constrained by the contentual part (including those reals that are not numerical equations as well as those that are!) in a way similar to that in which theory is to be constrained by observational data in the natural sciences.

This way of reading allows us to make sense of the previously quoted remark (Hilbert 1927:475) in which he compared theorizing in mathematics with theorizing in the natural sciences. There he stressed the idea that a theory "by its very nature" is supposed to eliminate the need "to fall back on intuition or meaning in the midst of an argument" and to allow argument to proceed "on the basis of a pure formula game." This is just another way of saying that what is essential to science is not contentual laws, but ideal propositions and derivations. Smorynski's fgps, being contentual laws, could not function as part of a "pure formula game." Therefore, it is doubtful that they are what Hilbert had in mind when he spoke of an essential, nonequational component of the science of mathematics.

Smorynski's first ground for introducing the "new" category of fgps is thus implausible. Still, it is better than his second one, which contends that the new category of fgps is needed in order to make sense of Hilbert's views concerning the difference in meaningfulness between fgps, on the one hand, and existential generalizations, on the other. As is well known, Hilbert emphasized the asymmetry of these two kinds of propositions, treating the former but not the latter as meaningful, an attitude which has often puzzled students of Hilbert (myself included). Smorynski claims that if one reads Hilbert as advocating a separate category (separate, that is, from the reals and ideals) for the fgps, then one can make perfect sense of this otherwise puzzling view. The argument given for this claim is the following (pp. 59–60).

> Real propositions do not exhaust the class of finitistic prop ositions. There are also what I shall call here the finitary general propositions—assertions of the form "for every numeral n, n+1=1+n", which Hilbert would have written in Leipzig as the free-variable formula,
>
> $$x + 1 = 1 + x.$$
>
> The general assertion "is from the finitist point of view *incapable of being negated*" because the negation ("for some numeral n…") "cannot be interpreted as a combination, formed by means of 'and', of infinitely many numerical equations". [It may seem odd that infinite conjunctions are finitistic assertions. However, i. one cannot seriously consider any science which does not propose universal laws, ii. there are finitary schematic methods of proof, iii. consistency, which must be proven finitistically, is such a universal assertion, and iv. it was his *existential* theorems that Hilbert had been criticized for.]

Anyone familiar with the passage from which Smorynski's paraphrase is wrested will immediately see that something has gone wrong. Hilbert did *not* view finitary generalizations as infinite conjunctions; indeed, he presented infinite conjunctions as quintessential cases of *nonfinitary* propositions. He therefore would not have used nonequivalence to an infinite conjunction as a standard for establishing the nonfinitary character of existential generalizations. What Smorynski has done, then, is to produce an interpretation of the "paraphrased" remarks that is not only different from, but, indeed, the very antithesis of, what Hilbert in fact said.

This mistaken interpretation of Hilbert is the result of a simple misreading of the relevant texts. To locate the mistake, we shall quote the passage from which Smorynski's paraphrase is taken, first the German original and then its standard English translation.

> Wir stoßen also hier auf das Transfinite durch Zerlegung einer existentialen Aussage, die sich nicht als eine OderVerknupfung deuten läßt. Desgleichen kommen wir zu transfiniten Aussagen, wenn wir eine allgemeine, d.h. auf beliebige Zahlzeichen sich erstreckende Behauptung negieren. So ist z. B. die Aussage, daß, wenn a ein Zahlzeichen ist, stets
>
> $$a + 1 = 1 + a$$
>
> sein muß, vom finiten Standpunkt *nicht negationsfähig*. Dies können wir uns klar machen, indem wir bedenken, daß diese Aussage nicht als eine Verbindung unendlich vieler Zahlgleichungen durch "und" gedeutet warden darf, sondern nur als ein hypothetisches Urteil, welches etwas behauptet für den Fall, daß ein Zahlzeichen vorliegt.

That is the German original from *Mathematische Annalen,* 95 (1926), 173. Now follows Bauer-Mengelberg's translation from van Heijenoort (1967:378), with which we agree.

> Thus we encounter the transfinite when from an existential proposition we extract a partial proposition that cannot be regarded as a disjunction. In like manner we come upon a transfinite proposition when we negate a universal assertion, that is, one that extends to arbitrary numerals. So, for example, the proposition that, if n is a numeral, we must always have
>
> $$n + 1 = 1 + n$$
>
> is from the finitist point of view *incapable of being negated.* This will become clear for us if we reflect upon the fact that [from this point of view] the proposition cannot be interpreted as a combination, formed by means of 'and', of infinitely many numerical equations, but only as a hypothetical judgement that comes to assert something when a numeral is given.

The crucial phrase is "diese Aussage nicht als eine Verbindung unendlich vieler Zahlgleichungen durch 'und' gedeutet werden darf (translated as "the proposition cannot be interpreted as a combination, formed by means of 'and,' of infinitely many numerical equations…," by Bauer-Mengelberg, although the "the" would probably better—if less mellifluously—be translated as "this"). And the crucial question regarding this phrase is this: to which "Aussage" (proposition) is Hilbert referring?

The answer, of course, is that he is referring to the proposition "for every numeral n, n+1=1+n," and what he is saying about it is that it cannot be interpreted as an infinite conjunction. This simple and obvious reading, however, is directly opposed to Smorynski's. He takes the "diese Aussage" of the passage quoted to refer not to "for every numeral n, n+1=1+n," but rather to its denial! Thus, according to Smorynski, what Hilbert was saying when he wrote "diese Aussage nicht als eine Verbindung unendlich vieler Zahlgleichungen durch 'und' gedeutet werden darf is that the unbounded existential proposition "for some numeral n, n+1 1+n" (i.e. the denial of the fgp "for every numeral n, n+1=1+n") cannot be interpreted as an infinite conjunction, and that it is on this account that it is nonfinitary in character.

Thus, like his first ground, Smorynski's second ground for introducing the fgps is mistaken. Hilbert's classification of universal generalizations as meaningful and (unbounded) existential generalizations as meaningless was not due to any belief on his part that the former but not the latter are infinite conjunctions. Hilbert did not take universal generalizations to be infinite conjunctions. Nor did he take interpretability as an infinite conjunction to be a criterion of finitary meaningfulness; indeed, he was of the contrary opinion.

Smorynski's chief mistake is a failure to properly recognize the ways in which Hilbert saw the real propositions as being subdivided. His basic—and most important—distinction was that of separating the *problematic* from the *unproblematic* reals. When a finitary proposition or expression is such that one cannot apply the full range of classical logical operations to it without generating a nonfinitary proposition, then it is what Hilbert broadly referred to as "problematic." But not all problematic finitary propositions are problematic for the same reasons. A bounded existential quantification like "there is a prime greater than **p** but less than **p**!+1" is a problematic real proposition because it classically implies the unbounded existential quantification "there is a prime greater than **p**,"which is not a finitary proposition (cf. Hilbert 1925:377–8).

Finitary generalizations, too, are problematic, but for a different reason. When a finitary generalization is negated, one does not get a finitary proposition at all (nor even a finitary propositionschema). One gets instead an ideal proposition (which is neither a genuine proposition nor a proposition-schema, but rather some sort of formula or computational device). Thus finitary generalizations are problematic because the classically valid law of excluded middle cannot freely be applied to them (cf. Hilbert 1925:378).[6]

This, in brief, is Hilbert's account of finitary generalizations. It shows what, in Hilbert's view, separates finitary generalizations from other finitary propositions, and why, contrary to what Smorynski suggests, they are to be seen as forming a subcategory of the reals rather than a separate category alongside them. It does not, however, say what is supposed to separate finitary generalizations from ideal generalizations. We must therefore briefly address this question.

On Hilbert's view what sets the two apart is the different ways in which they are instantiated. Part of this difference consists in the different character of the instances themselves; or, rather, the differences between the expressions substituted for the variables in producing instantiations of finitary generalizations and ideal generalizations. Hilbert characterized this difference as follows (cf. Hilbert 1931:194):

> Es sei hier daran errinert, daß die Aussage (x)F(x) viel weiter reicht als die Formel F(z), wo z eine beliebig vorgelegte Ziffer ist. Denn im ersteren Falle darf in F(x) für x nicht bloß eine Ziffer, sondern auch ein jeder in unserem Formalismus gebildete Ausdrück vom Zahlcharakter eingesetzt werden.[7]

Smorynski presents this as the whole of Hilbert's distinction between finitary and ideal generalizations, saying that the difference between the two is that the former take only numerals (and other finitarily well-defined terms) as substituends of the variables, while the latter also admit "meaningless infinitary constructs" as substituends (cf. Smorynski 1988:63–4). But though this is part of the difference between finitary generalizations and ideal propositions, it is not the whole difference, nor, as we shall now attempt to show, even the major difference.

Hilbert repeatedly stressed the fact that there are formulas like "1+*x*=*x*+1" which, though perhaps *interpretable as* finitary generalizations, are nonetheless ideal propositions. One example of this (cf. 1925:379–80) is the formula *"x+y=y+ x"* This formula, he says, is admissible from the finitist point of view as expressing the generalization that for all (finitarily well-defined) numerals *m, n, m+n=n+m.* Yet despite this fact, he says, *"x+y=y+x"* is still an ideal proposition.

His reasons for saying this are important both for understanding the nature of ideal propositions and for seeing what it is that essentially separates them from finitary propositions (or proposition-schemata). Hilbert focused on the question of what *governs the use of* a given formula or proposition; specifically, whether it is the rules of some formal procedure (a "proof pro cedure," in his terminology), on the one hand, or consideration of the content of the proposition expressed by the formula in question, on the other. An ideal formula, he says, is not such that its use is guided by appeal to the content of any proposition that the formula in question might (under an assumed semantics) express. Rather it is governed by a given system of rules of "algebraic" or syntactical manipulation. The use of a finitary formula (or proposition-schema), on the other hand, *is* guided by the considerations of the content (including its evidentness) of the proposition which it (under a given finitary interpretation) expresses.

It follows from this that what determines whether a formula like "1+*x*=*x*+1" is finitary or ideal is not whether, under a specified interpretational scheme, it expresses a finitary proposition or proposition-schema. Nor is it whether the substituends for the variables of the formula are restricted to the class of finitarily well-defined terms. Rather, it is whether the *use* of the formula is governed by the rules of a formal "proof procedure," or by considerations of the content of the proposition which it expresses. This seems to be what Hilbert had in mind when he insisted that (cf. 1925:380)

> even when a proposition, so long as it is combined with some indication as to its contentual interpretation, is still admissible from our finitist point of view, as, for example, the proposition that always
> $$m + n = n + m$$
> where m and n stand for specific numerals, we do not select this form of communication but rather take the formula
> $$a + b = b + a.$$
> This is no longer an immediate communication of something contentual at all, but a certain formal object, which is related to the original finitary propositions
> $$2 + 3 = 3 + 2$$
> and
> $$5 + 7 = 7 + 5$$
> by the fact that, if we substitute numerals 2, 3, 5, and 7, for a and b in that formula (that is, if we employ a proof procedure, albeit a very simple one), we obtain these finitary particular propositions. Thus we arrive at the conception that a, b, =, and +, as well as the entire formula
> $$a + b = b + a$$
> do not mean anything in themselves, any more than numerals do. But from that formula we can indeed derive others; to those we ascribe a meaning, by treating them as communications of finitary propositions. If we generalize this conception, mathematics becomes an inventory of formulas—first, formulas to which contentual communications of finitary propositions [hence, in the main, numerical equations and inequalities] correspond and, second, further formulas that mean nothing in themselves and are the *ideal objects of our theory.*[8]

In short, then, a formal proof procedure whose substitutions are restricted to those in which a variable-occurrence is replaced by a numeral or other finitarily well-defined term is, for all that, still a formal proof procedure. And that is why "1+*x*=*x*+1" functions as an ideal formula when one obtains "1+3=3+1" from it by appealing to a simple proof procedure which calls for substitution of the numeral "3" for the variable *"x"* (cf. the passage from Hilbert (1927) quoted in note 8). If the use of a formula is governed by such a procedure, then it functions (in that context) as an ideal formula, regardless of whether the substituends for its variables are finitarily well-defined. If, on the other hand, a formula like "1+*x*=*x*+1" is used to formalize a piece of reasoning that is based on considerations of the content of the proposition it expresses, then, in Hilbert's view, it functions as a real formula. The critical question is thus whether a formula is used to formalize a proposition which plays a role in a contentual argument, or whether it is used sheerly as an element in a formal "computational" procedure.

(Actually, things are more complicated than even the above remarks would indicate. For Hilbert may have assumed that when we employ a formula like "1+*x*=*x*+1" in an ideal way (i.e. as part of a proof procedure), its logic (syntactically

speaking) is classical logic. If this is so, then, since classical quantification is quantification over *objects* rather than *objects-as-referred-to-in some-particular-way,* the substitution class for *"x"* would have to be the wider class including terms which are not finitarily welldefined. Perhaps Hilbert had some such assumption in mind; at any rate, it seems to fit with his idea that the point of having ideal methods in mathematics is to preserve the natural and efficient rules of classical logic.)

Hilbert's real/ideal distinction thus consisted of a major division between the real and the ideal propositions, and a minor (sub)division of the real propositions into problematic and unproblematic. In this scheme, the finitary generalizations are merely one particular type of problematic real proposition and not, as in Smorynski's account, a separate class of elements standing alongside the reals and the ideals, nor even a third subcategory of the reals standing alongside the subcategories of problematic and unproblematic. Failing to see this, Smorynski misses the basic point of Hilbert's subdivision of finitary thought; namely, that of separating those finitary elements whose logic abides by the "natural" principles of classical reasoning (the so-called "unproblematic" reals) from those (the problematic) whose logic is the more cumbersome and less "natural" one of finitary reasoning. Missing this basic point, he thus confuses both the general character of Hilbert's conception and its underlying motivation.

III.
THE G1-BASED ARGUMENT AGAINSTHILBERT'S PROGRAM

With the above discussion of the real/ideal distinction as background, we may now address the main concern of this paper, namely, the G1-based argument against HP. Smorynski (1985: 3–4) presents this argument as follows.

> Hilbert's Programme can be described thus: There are two systems, nowadays called formal theories, S and T of mathematics. S consists of the finite, meaningful statements and methods of proof and T the transfinite, idealized such statements and methods. The goal is to show that, for any meaningful assertion G, if T ⊢G then S ⊢G. Moreover, this is to be shown in the system S.
>
> Gödel destroyed Hilbert's Programme with his First Incompleteness Theorem by which he produced a sentence satisfying a sufficiently narrow criterion of meaningfulness and which, though readily recognized as true—hence a theorem of the transfinite system T, was unprovable in S. In short, he produced a direct counterexample to Hilbert's desired conservation result.

In order to properly evaluate this argument, it will prove useful to give a more explicit version of it. As a step in this direction, we may begin by noting that, in Smorynski's argument, the ideal theory T is treated as representing a *norm* for ideal theorizing; that is, it is taken to be an ideal theory which proves those sentences that an ideal theory *ought* to prove. These real sentences, on Smorynski's view, are those which are "readily recognized as true." The norm in operation here is thus a type of completeness constraint requiring that an ideal theory be complete with respect to those real sentences in its language that are readily recognizable as true.

The assumption that T has this normative status is critically important to the success of Smorynski's argument. This is so because (i) it is only in its capacity as a prover of the recognizedly true real sentences that T can be said to prove the Gödel sentence G, (ii) it is only in its capacity as a prover of G that T can be shown (by appeal to G1 for S) to prove a real sentence not provable in S,[9] and (iii) it is only because T proves some real sentence not provable in S that it fails to be conservative and hence subject to the sort of defense to which Smorynski takes HP to be committed. The basic thrust of Smorynski's argument, then, is that an ideal theory is deficient if it fails to prove every recognizable real truth (formulable in its language) as a theorem, and that a deficiency of this sort would be roughly as serious as a violation of conservation. Thus, ultimately, the G1-based argument is intended to present a dilemma to the following effect: the Hilbertian's ideal theories must either fail in their obligation to prove all true real sentences (formulable in their respective languages), or they must fail to be conservative with respect to real mathematics. Either way, HP fails.

We find this argument unconvincing for two reasons which we shall now briefly sketch and discuss in greater detail below. The first of these concerns the assumption (hereinafter referred to as *Smorynski-completeness*) that, to be adequate, an ideal theory must prove all recognizedly true real sentences of its language. On an instrumentalist conception of ideal mathematics (which is what we take the conception adopted by the Hilbertian to be), the goal of ideal theorizing is the relatively modest one of proving more efficiently what real theorizing would only prove less efficiently. Specifically, there is no basis in the Hilbertian's instrumentalistic program for requiring that the ideal theory prove *more* real results than the real theory that it is intended to replace. This being so, there is no evident need for the ideal theory T to prove a given sentence unless S also proves it. As shall be argued below, this feature of the Hilbertian's program calls the legitimacy of Smorynski-completeness into doubt.

Our second concern regarding the G1-based argument centers on its demand (henceforth referred to as the *conservation condition*) that ideal mathematics be a conservative extension of real mathematics. The likening of the role of real proof to that

of observational verification in the natural sciences suggests that the basic constraint on ideal theorizing is that of a *soundness condition* requiring that ideal mathematics not prove any real theorem that is refutable by real means. But whereas such a soundness condition naturally gives rise to a corresponding conservation condition in the case of observation statements, it does not do so in the case of real propositions. Hence, the conservation condition is called into doubt.

These then, in outline, are our two main objections to the G1-based argument. We shall develop each in greater detail below. Each, we believe, is sufficient by itself to refute the G1-based argument. However, our concern is not so much to refute that argument as to understand it and its limitations. Since both objections promise to contribute something to this, both would seem to be worthy of the more detailed discussion we now give them.

In order for a theory belonging to ideal mathematics to be adequate, must it be complete with respect to the true real sentences formulable in its language?

In order to answer this question, we must begin by clarifying what is meant by "true real Sentence," and our claim is that it is a classical rather than a constructive (specifically, a finitaryconstructive) notion of truth that is involved in here. Thus, what Smorynski-completeness demands is that all *classically* true real sentences formulable in L(T) also be provable in T.

That this is the way that Smorynski-completeness should be understood is apparent from the fact that the notion of truth appearing in it must be the same as that according to which G is true, and that notion of truth is the classical one. The intent of Smorynski-completeness is clearly to lay down a requirement on ideal mathematics which T fails to satisfy by failing to prove G. But what makes G true is the truth of its instances, not its provability by finitary means. Hence, it is only classically and not constructively (in particular, not finitarily) true. Consequently, T's "failure" to prove G can only be counted as a "failure" to prove all classically true real sentences formulable in its language; which means that Smorynski-completeness must be seen as the requirement that an ideal theory T prove all classically true real sentences formulable in its language.

The significance of this fact for our argument is that it clarifies the possible defenses for Smorynski-completeness as a constraint on ideal theorizing. In particular, it shows us that it cannot merely be seen as an attempt to enforce a simple *strength* requirement on T to the effect that T be powerful enough to codify the whole of finitary reasoning. This is clear from the fact that the theorems of finitary reasoning are not at all the same as the classically true real sentences formulable in L(T), since the latter clearly include sentences that do not belong to the former. What we must now consider is whether there is some other justification for it.

We believe that there is not, and that this is clear from the instrumentalist character of HP. Smorynski-completeness is attractive to a realist, since one of the realist's primary duties in constructing a theory is to bring all truths pertaining to its subject matter under its purview. But the responsibilities of the instrumentalist are different. His task is to replace a less efficient means for identifying a given body of truths with a more efficient one. But that creates no obligation for him to prove anything more than is provable by the more cumbersome methods that he would replace. Once this point is properly appreciated, the inappropriateness of Smorynski-completeness as a criterion of adequacy for ideal theorizing is apparent. Since G is not provable by real means, there is no reason why an ideal theory—whose aim is to improve upon the efficiency of real methods of proof—should be required to prove G either.

Should ideal mathematics be a conservative extension of real mathematics?

The central idea of the argument of the preceding subsection can readily be extended to provide the starting point for the argument of this one. This extended idea runs as follows: since (i) the Hilbertian's chief obligation is to the efficiency and reliability of his ideal systems, and since (ii) there is no apparent connection between these properties and the ability of an ideal system to decide every real proposition formulable in its language (this being suggested especially by the fact that finitary reasoning does not itself decide every real sentence), it follows that (iii) there is like-wise no apparent reason to demand that an ideal theory decide every real proposition formulable in its language. The significance of this conclusion for the present argument is as follows: without a requirement of the finitary decidability of real sentences, the usual conservation condition cannot be derived from the more basic requirement of soundness in the way typical of scientific theories generally.

(Demanding that an ideal theory T decide the same range of real sentences as is decided by its real counterpart S is not at all the same thing as demanding that T decide all real sentences formulable in L(T). This is so because S itself might not, and indeed typically does not, decide all real sentences formulable in L(T). The Gödel sentence G is a good example. Thus, T may have the scope necessary to serve as a replacement for S (i.e. T may decide every real sentence decidable by S) even if it does not decide every real proposition formulable in its language.)

We begin our argument by noting that the most important feature of S (=real mathematics) is its (presumed) epistemic authority. By this we mean that it is supposed to be the final judge concerning the truth or falsity of real sentences. Its veracity is thus treated as assured, which means that any real proposition it decides must be decided in the same way by any ideal theory which also decides it.

This condition on the real results of ideal reasoning does not, however, imply the conservation condition. It is one thing to say that any real proposition   which S decides must be decided in the same way by T, if T decides it, and quite another to say that S must prove every real proposition   provable in T. The reason for this divergence is, of course, that S may not decide   at all; in which case, respect for the epistemic authority of S provides no reason to demand that   be provable in T only if it is also provable in S. To put it another way, when S does not decide   T cannot transgress against the authority of S by deciding
   Thus, under such circumstances it is freed from the obligation to adhere to the conservation condition. We conclude, therefore, that the conservation condition is too strong, and should be replaced by the following weaker condition.

(Weak Conservation)   For every real sentence r of L(T) such that r is decided by S, if r is provable in T, then r is provable in S.[10]

If one's account of real sentences and finitary reasoning allows that there are real sentences that are not finitarily decidable (i.e. not S-decidable), then the weaker condition just stated entails no obligation to prove of those real sentences that they are provable in T only if provable in S. Thus, in particular, it does not entail an obligation to prove that G is provable in T only if it is provable in S. Nor, indeed, does it require that this even be true. The unprovability of G in T therefore does not constitute a counterexample to Weak Conservation. Consequently, even if (contrary to the argument of the preceding subsection) T were required to prove G in order to be an adequate ideal theory, G's unprovability in S would not constitute a violation of the more appropriate condition of Weak Conservation.
   To allay the suspicions of those who are drawn to the usual conservation condition, it is necessary to offer some explanation of why, despite its ultimate indefensibility, the ordinary conservation condition should nonetheless prove so alluring. Our attempt at doing so begins with a very basic tenet of Hilbert's Program; namely, the alleged analogy, mentioned earlier, between the epistemic roles played by observation in physics and real proof in mathematics. This analogy suggests, albeit falsely, that if conservation with respect to observational consequences is a reasonable condition to place on a physical theory,[11] then conservation with respect to real consequences ought to be a reasonable condition to place on an ideal mathematical theory. Thus, the reasoning continues, if it can be shown that conservation with respect to observational consequences is a reasonable condition to place on physical theorizing, it should follow that conservation with respect to real consequences is a reasonable condition to place on the Hilbertian's ideal theorizing.
   One can, of course, argue quite convincingly that conservation with respect to observational consequences is a reasonable condition to place on physical theorizing. The argument begins by noting that in physics (as in natural science generally) observation is granted a place of special epistemic privilege; by which is meant that the theoretical elements of physics are not at liberty to oppose that which is established by observational means. This epistemic subordination of theoretical proof of observational verification is expressed as a principle of *observational soundness,* which is then taken to constitute a condition of adequacy on any body T of theoretical reasoning in physics:

(OS)   For any observation sentence O, if T proves O, then O cannot be observationally falsified.

From (OS) as a starting point, it is possible to argue for the following principle of *observational conservation,*

(OC)   For any observation sentence O, if T proves O, then O is verifiable by observational means,

by adding to (OS) the further premise that observation sentences are (by their very nature?) supposed to be decidable by observational evidence. If, as (OS) requires, no observational consequence O of T can be observationally refuted, and every observation sentence is observationally decidable, then every observational consequence of T must be observationally provable. Hence, (OC) follows from (OS) and inherits its plausibility.
   Can a similar defense for a condition of real-conservation on ideal theorizing be given? A condition of *real-soundness* would appear to be no less defensible a constraint on ideal theorizing than (OS) is on physical theorizing:

(RS)   For any real sentence r, if T proves r, then r cannot be refuted by real means (i.e. ¬r is not provable in S).

But in order to get a condition of *real-conservation* from (RS), one requires an additional premise stating that every real sentence is decidable in S; and such a premise is false—at least if one counts such sentences as G as real.[12] It is therefore impossible to base a case for real-conservation on the presumed need for real-soundness in the same way that a case for observational complete ness can be based on a presumption of need for observational soundness.
   We conclude our discussion of the conservation condition by briefly considering a possible objection to the argument just given. This objection focuses on that feature of real sentences which is responsible for the above-noted separation of real-conservation from real-soundness, namely, their undecidability by real means. It has, moreover, both a conceptual and a

historical side to it. On the conceptual side, it holds to a constructivistic understanding of real propositions (including those generalizations which Hilbert referred to as "hypothetical judgements") and maintains that such a view demands that genuinely real propositions be decidable by finitary means. On the historical side, it would maintain that Hilbert was originally committed to the finitary decidability of real propositions and present G1 as having refuted that belief. Hence, it views Hilbert's original Program as having based a commitment to real-conservation on a deeper commitment to real-soundness, and therefore sees G1 as destroying that hope.

On the textual-historical side, there is the well-known remark in the "Mathematical Problems" address claiming that "every definite mathematical problem must necessarily be susceptible of an exact settlement, either in the form of an actual answer to the question asked, or by a proof of the impossibility of its solution and there-with the necessary failure of all attempts [to solve it]" (1901:444, brackets added). This same theme was sounded in Hilbert (1925), in the famous claim stating that in mathematics there is no *ignoramibus.* If "definite problems" are just real propositions in interrogative form, it is hard to see how Hilbert could have counted undecidability proofs as settling genuine problems while also holding the position that every real proposition must be finitarily decidable. Since propositions stating the undecidability of a given proposition by various means were held by Hilbert to be genuine propositions, this suggests that he would not have accepted the view that all genuine (=real) propositions must be finitarily decidable.

These textual points aside, however, it is doubtful that a constructivist semantics for real propositions *should* equate meaningfulness with finitary decidability. On a constructivist account, a user of the language is said to know the meaning of a real sentence r when for every proof (respectively refutation) of r, she *would* recognize as such *were* she to be presented with it. This does not imply, however, that real (i.e. meaningful) sentences are finitarily decidable—not even if it is assumed that the only kinds of proofs or refutations of real sentences there are are finitary. That this is so follows from the fact that being in a position to recognize a proof or refutation of r were one to be presented with it does not require knowing how to prove or refute it. Ability to tell of a given proof whether or not it is a proof or refutation of r does not imply ability to actually generate such a proof or refutation. There is thus no reason to maintain that, on a constructivistic account of their meaning, real sentences must be finitarily decidable.

Since, then, there are textual reasons for denying that Hilbert held the view that real propositions are finitarily decidable, and conceptual reasons against such a view regardless of its textual pedigree, it seems only charitable to withhold attributing such a view to Hilbert. We therefore reject the suggestion that he was committed to the finitary decidability of real propositions and that his commitment to real-soundness should therefore be seen as engendering a commitment to real-conservation.[13]

IV.
CONCLUSIONS

The argument of the preceding section defeats the G1-based argument against HP. In so doing, it re-focuses attention on the G2-based argument. This, we believe, is all to the good, since it calls attention to those problems which represent the deepest philosophical issues concerning Gödel's theorems and their implications for Hilbert's Program, issues on which the literature including the philosophical literature—has been strangely silent. At bottom, the central question is "What is an instrumentalist theory?" We shall argue that the proof of G2 (in particular, the familiar Derivability Conditions) places constraints on what is to count as an ideal theory that are unwarrantedly strict. Moreover, we shall argue that the G2-based argument against HP fails to properly distinguish two very different concerns regarding instrumentalist theories; namely, (i) whether they prove all they need to prove in order to be adequate replacements for the real theories that they are supposed to replace, and (ii) whether they prove only such real theorems as are not refutable by real means. Specifically, it fails to take proper account of the fact that the Hilbertian instrumentalist requires a finitary proof only of (ii) and not of (i), and that the two should therefore not be merged into a single condition to be finitarily proven.

Once a suitably liberal standard of what is to count as an instrumentalist theory is adopted, and once it is realized that, though the Hilbertian may have need of a finitary proof of the real-soundness of his ideal theories, he has no similar need for finitary assurance that they prove all of what he wants them to prove, the force of the G2-based argument is dissipated. Such, at any rate, is our thesis, though our basic aim is not so much to establish it as to call attention to those deeper philosophical issues from which it arises.[14]

These issues can perhaps best be got at by considering the differences between the conditions which the proofs of G1 and G2 place on their underlying notions of theory. The proof of G1 makes only minimal demands. It is concerned only with the content of a theory; i.e. with *what* it proves. Specifically, it demands only that the theory in question contain enough recursive number theory to weakly represent the set of (Gödel numbers of) its theorems.[15] One might therefore say that the notion of theory presupposed by the proof of G1 is one which identifies a theory with the set of theorems that it proves, and whose only additional constraint is that it contain enough recursive number theory to represent that set.

G2, on the other hand, makes more extensive demands on the notion of theory. Its proof depends not only on a theory's having the right content, but also on that content's having been admitted as such by a certain type of procedure. Two theories T_1 and T_2 may contain exactly the same entities in a given metamathematical category (e.g. the category of formulas, axioms, proofs, or theorems) and yet differ significantly with respect to the conditions used to quantify items for membership in it.

To illustrate what we are talking about, let us recall the examples of consistency-minded theories mentioned in the introduction and how they contrast with the standard conception of a formal axiomatic theory. On the standard conception, the basic metamathematical categories pertaining to a theory (e.g. those of formulas and axioms) are defined inductively, and the remaining categories (e.g. that of the theorems) are defined in terms of these. One begins the definition of a basic category by exhibiting a finite set of basic members, and then identifying the remaining members with those items that can be generated from the basic members (perhaps taken in combination) by applying (in some instances iteratedly) certain specified syntactical operations to them. The application of these operations is, moreover, supposed to be selfcontained in a certain sense. Specifically, it is supposed that in order to determine of a given item whether it belongs to the category in question, one need only attend to the syntactical traits of *its* constituents and their arrangement and not (even potentially) to the syntactical characteristics of other items.

Consistency-minded theories are different. They begin with metamathematical definitions of the standard sort, but then add various conditions which, broadly speaking, are consistency conditions of one or another sort. Thus, to take the Rosserian variant of consistency-minded theories as an example, one begins with a theory T, defined in the standard way, and forms the Rosser variant T_R of it by adding to the standard definition of proof a constraint to the effect that the last line of a proof-in-T_R not be the contradiction of the last line of any proof-in-T which precedes it in a given   -ordering of the proofs-in-T.[16] Thus, the category of proofs-in-T_R is generated by a different testing or qualifying procedure than the category of proofs-in-T, and that is what we mean by saying that the category of proofs-in-T is generated or governed by a different admission procedure than the (possibly coextensive) category of proofs-in-T_R.

Such differences separating the admission procedures for the various metamathematical categories of T from those of T_R will not have any effect on whether G1 holds for them. It will hold either of both or of neither, depending on whether T proves enough number theory to weakly represent the class of (Gödel numbers of) its theorems. Such differences, however, can have a decisive effect on whether G2 applies equally to both T and T_R, since the proof for G2 is sensitive to the character of the admission procedures for the various metamathematical categories pertaining to a given theory.

What this means is that G2 applies not to sets of theorems (or even sets of proofs!) but rather to sets of theorems-as-admittedby-a-particular-type-of-procedure. This raises the possibility—of crucial importance to the present discussion—that there might be nonstandard ways of generating sets of theorems and proofs that are *not* covered by the proof of G2, but which nonetheless constitute perfectly sensible conceptions of theory for the Hilbertian instrumentalist. Were this to be so, the Hilbertian would be free to construct ideal theories not ruled out by G2, and having the same content as those that are. At any rate, this would seem to be a possibility worth looking into.

These last remarks point out why it is so important to get clear about the success of the G1-based argument against HP. For if the G1-based argument were correct, then the prohibition against the Hilbertian would be one bounding the strength or content of his ideal theories (i.e. one fixing limits on what his ideal theories could be capable of proving) and not just one restricting the mode according to which that content is to be generated. If, on the other hand, the G1-based argument is not successful, and the basis of evaluation for HP must accordingly be shifted to G2, then the focus of concern is not one concerning strength or content, but rather one concerning mode of generation (with respect to which the Hilbertian may well have flexibilities that he does not have with respect to the strength of his ideal theories). We regard this as a difference of potentially great significance, and one which makes the investigation of consistency-minded modes of theorizing imperative.[17]

Perhaps the most basic issue raised by the possibility of consistency-minded theorizing is this. Are theories to be viewed *extensionally* (i.e. as sets of beliefs) or *intensionally* (i.e. as methods or procedures for selecting beliefs)? The good theory, of course, both targets particular propositions for belief and does so by bringing them under a method which testifies to their beliefworthiness. Let us refer to the first of these two components of good theorizing as the *locative component* and the second as the *methodological component.*

Neither component, by itself, is sufficient for good theorizing. A good theory is not a mere "list" of propositions to believe, with no credential provided to attest to their belief-worthiness. Likewise, a good theory is more than sheer method with no means given for finding a set of particular propositions that are to be believed.[18]

All this may seem so elementary as to scarcely be worth mentioning. Why then do we emphasize it so? The answer is that it points to an important yet easily overlooked truth concerning the Hilbertian's project; namely, that the locative and methodological features of theorizing are *separable.* In order to defend a given body of ideal theorizing, the Hilbertian must know something about both its locative and its methodological features. But what he must know about them—and this is the important point—is different. Regarding the methodological element, what he must know is that it is real-sound (i.e. that all of the real sentences it decides are decided in the same way by finitary means). Furthermore, this must be made apparent by

finitary means if his use of ideal method is to avoid the threat of "diluting" the knowledge that he might otherwise obtain by forswearing the use of ideal methods and sticking to (the presumably less efficient) real methods. A gain in efficiency is not so attractive if it brings with it a corresponding drop in the epistemic quality of the more efficiently attained "knowledge." The Hilbertian thus proposes to replace the object-level real proofs of a given real sentence   with a meta-level real proof of that is composed of two elements: (i) a real metamathematical proof that   is provable in the ideal system T (which would consist in producing a proof of   in T), and (ii) a real metamathematical proof showing of   that, if it is provable in T, then it is also provable by real means at the object-level.[19]

Having this sort of control over the quality of his ideal methods assures the Hilbertian instrumentalist that their use will not engender "dilution" (i.e. will not result in an epistemic product of a quality lower than what could have been attained by sticking to the real methods that the ideal methods in question are intended to replace), and this is what he needs to know.[20] What the Hilbertian demands of the locative component of a given ideal theory T is that it replicate the locative capacity of that body R of less efficient real reasoning which it is intended to replace. In other words, the Hilbertian must be able to show that every sentence provable in R is also provable in T.[21] However, the Hilbertian has no need for *finitary assurance* that a proposed ideal replacement extensionally simulates the real reasoning it is to replace. Hence, he is under no obligation to prove finitarily that each theorem of R is a theorem of T.

This is significant because though the move from a standard ideal theory T to one of its consistency-minded variants T generally affords a greater measure of control over the quality of its real theorems, this increase in methodological control brings with it a consequent loss in ability to prove (by means codifiable in T, and hence, typically, by finitary means) the extensional equivalence of T and T '.[22] But, as the reasoning of the preceding paragraph suggests, this "loss" is not disabling for the Hilbertian. His primary obligation with regard to the content or locative component of T is to show that it replicates R, not T. And even if he were able to do this only by first proving the extensional equival ence of T and T, all that would follow from his inability to prove the latter fact finitarily is his subsequent inability to *finitarily* show that T replicates R. But this is no impediment to his program since what he needs a *finitary* proof of is not the replication of R by T, but rather the Weak Conservation with respect to R of T. With respect to the replication of R by T, all he needs is convincing evidence, not finitary proof.[23]

Thus, the general fact that one cannot finitarily prove the theorem-wise equivalence of consistency-minded theories and their standard counterparts gives no reason why the Hilbertian should not use consistency-minded construction techniques for his ideal theories. It poses no obstacle to his getting the kind of control over the locative element of theorizing that he needs, and, since the usual techniques are susceptible to G2 in a way that the consistency-minded techniques are not, it also affords him an advantage over the standard techniques when it comes to managing the methodological factor. This does not, however, show— what is a very difficult question, and one which we are currently not in a position to resolve[24]—that moving to consistency-minded techniques of theory construction will actually allow one to carry out HP. Nonetheless, it does suggest that part of the traditionally pessimistic view of the Hilbertian's prospects is due to an unwarrantedly narrow conception of how ideal theories should be constructed, a conception which is built into the very fabric of the Derivability Conditions governing G2, and which systematically ignores the possible benefits to the Hilbertian of adopting consistency-minded modes of ideal theorizing.

NOTES

1 Cf. Hilbert (1927:475) for an explicit statement of the analogy between physical theory and ideal mathematics. Smorynski too (cf. 1988:59) adopts this view. Thus he quotes the remark from Hilbert (1925) that the science of mathematics is not reducible to its real elements, but that it must always yield correct real results, and then goes on to say that one can see in this remark the view that "Mathematics is an abstract theoretical science subject to numerical control just as physics is an abstract theoretical science subject to experimental control."

2 Throughout this paper, when we speak of an ideal proof we shall generally mean an ideal proof of a real sentence.

3 In noting that, on the instrumentalist conception, ideal proofs "yield" (via evaluation in a real metamathematical scheme), rather than constitute, justifications, the question is naturally raised as to why they should be of any epistemic interest. Why, that is, should one be interested in producing justification through the indirect means of ideal proof rather than through the direct means of real proof? Such questions, however, apply as well to the realist conception; for one can as well ask "Why should we be interested in pursuing justification for observational statements through the indirect means of theoretical arguments rather than through the more direct means of observation?"

   In the empirical sciences, the question is perhaps a little easier to answer than in mathematics. For there, at least sometimes, the whole idea is to not be in a position to observationally settle a question (e.g. when it is the empirical effects at ground-zero of a

nuclear explosion that are in question). Other times, it might not be disutility, but rather possibility, feasibility, cost, and/or inconvenience that would make an alternative to observational justification attractive. In any event, there must be something about justification via use of ideal methods (be they instrumentalistically or realistically conceived) that makes it attractive as an alternative to real justification. In Hilbert's case this has to do with what might broadly be termed "efficiency." On his view, there is a difference between the laws according to which our minds operate most efficiently (namely the laws of classical logic), and the laws according to which finitary (i.e. real) truth works (i.e. finitistic logic). This is the epistemological predicament of the human who seeks mathematical knowledge. Hilbert's project was to provide a way out of the predicament by showing that we can enjoy the benefits of efficiency of (ideal) classical reasoning without sacrificing the accuracy of (real) finitary reasoning.

4   There are those (e.g. Gentzen 1936, 1938; Prawitz 1981) who have opted for a more realistic interpretation.

5   On p. 64, Smorynski says that, ignoring subtleties, "we can say that Hilbert said the following: Let S be a formal system of finitary arithmetic and let T be some system of transfinite mathematics. Suppose S proves the consistency of T.... Then: For any universal assertion G, if ⊢T G then ⊢S G." The latter is the conservation condition referred to in the text. It makes finitary general propositions function like data or controls on ideal (i.e. transfinite) mathematics, because it forces every finitary general proposition proven in the ideal theory to be corroborated by finitary means.

6   The interested reader may consult Detlefsen (1986: chs I, II) for further discussion of Hilbert's views of problematic reals (and some of the difficulties associated with them).

7   My translation: "It should be remembered here that the proposition *(x)F(x)* extends much farther than the formula *F(z),* where *z* may be any one of the specified numerals. For in the first place, one is permitted to substitute for *x* in *F(x)* not just a numeral, but any one of the expressions in our formalism that is constructed from numerical terms...."

   (Note that here *"(x)F(x)"* plays the part of an ideal generalization, and *"Fz"* the part of a finitary generalization.)

8   In this quotation, I have used the Latin characters "m," "n," where Hilbert used Gothic characters. The same general view is presented in Hilbert (1927:469–70), and there his example is exactly the same formula that Smorynski uses (though he writes it as "1 +a=a+1" instead of "1+x=x+1").

> algebra already goes considerably beyond contentual number theory. Even the formula
> $$1 + a = a + 1,$$
> for example, in which a is a genuine number-theoretic variable, in algebra no longer merely imparts information about something contentual but is a certain formal object, a provable formula, which in itself means nothing and whose proof cannot be based on content but requires appeal to the induction axiom.
> The formulas
> $$1 + 3 = 3 + 1 \text{ and } 1 + 7 = 7 + 1,$$
> which can be verified by contentual considerations, can be obtained from the algebraic formula above only by a proof procedure, such as formal substitution of the numerals 3 and 7 for a, that is, by the use of a rule of substitution.

9   Since Smorynski's argument depends upon T's proving G, it is clear that G cannot be the Gödel sentence for T. What may be less clear is what theory (or theories) it is of which G is supposed to be the Gödel sentence. As we shall see in a moment, the key constraints on G are that it be (classically) true and not provable in S. Thus, if we take S to stand for a formalization of finitary number theory, G might be taken to be the Gödel sentence of S. For taken in that way, S is true and hence consistent. Hence, its Gödel sentence is not provable in it. Hence, its Gödel sentence is (classically) true. This makes it clear, however, that G might just as well be taken to be the Gödel sentence of any extension of finitary number theory that is clearly consistent.

10 Stated a little more formally, this conservation condition reads as follows (where T is the ideal system whose conservation is to be proven, S the formalism representing finitary reasoning, and "⊢T" and "⊢S" stand for provability in T and S respectively):

   For any real sentence   of L(T) such that   is decidable in S,

11 Here, in speaking of a physical theory that is conservative with respect to its observational consequences, we mean a physical theory which is such that every observation statement it implies is also verifiable by observational means.

12 Nor is a gambit like Smorynski's distinction between real and finitary general propositions of any use here. What Smorynski (1988) calls "real sentences" are all decidable by finitary means. What he classifies as "finitary general propositions," however, are not, since G is treated as a finitary general proposition. However, the conservation condition is supposed to apply to finitary general propositions (cf. the formulation of conservation on p. 64 of Smorynski 1988) as well as to real propositions. Such a conservation condition is no more derivable from a corresponding soundness principle (i.e. a principle of soundness extending to Smorynski's finitary general propositions as well as to what he calls the real propositions) than (RC) is derivable from (RS). Hence, our argument would not be undone by employing Smorynski's scheme of distinctions.

13 This, of course, implies neither that Hilbert did not hold that the real sentences were, as a matter of contingent fact, finitarily decidable (though, as indicated by the earlier textual remarks, it is questionable that he did), nor that he was not surprised by Gödel's proof. It is to say only that such a belief, if held at all by Hilbert, was nonetheless extraneous to his program and not one of its essential tenets.

14 Detlefsen (1986:120–4) contains the only sustained argument for the adequacy of consistency-minded theories (in particular, Rosser variants) as models of instrumentalist theorizing. Aside from that, the only indications of awareness of the possible inappropriateness of the view of theory that is presupposed by the Derivability Conditions are in Kreisel and Takeuti (1974), Jeroslow (1975), and Kreisel (1980). The former (pp. 47–8) write that the consistency-minded variant of Rosser provides "a neat model for Wittgenstein's speculations" on the nature of rule-following. Jeroslow treats the consistency-minded theories of Feferman (1960) as based on a trial-

and-error conception of theory construction. Finally, Kreisel (1980:173) contains the statement that Rosser variants "mirror quite well, albeit crudely, an essential method used in practice for checking proofs: *comparison with background knowledge…"*

15 A set of numbers   is said to be weakly representable in T if there is a formula   of L(T) of one free variable such that for every natural number (where *"n"* is the standard term in L(T) for *n*).

16 When we speak of the Rosser system T_R "starting with" a standardly defined system T, we are supposing that the standardly defined system T serves as a specification of a body of standard ideal derivations whose advantages we should somehow like to incorporate into our consistency-minded replacement T_R of R. It should thus be clear that there is nothing sacrosanct about building consistency-minded theories from standard ones, and that we would only do so in those cases where we were convinced of their utility as ideal instruments.

17 How serious are the bounds on the strength of the Hilbertian's ideal theories that are induced by Gödel's work? The so-called "reverse mathematics" of Friedman and Simpson attempts to show that it is not as bad as might be thought. They have proven that there are systems embodying a substantial portion of classical mathematics all of whose $_2$ theorems are provable in PRA; and Sieg (1985) has subsequently shown how to prove this by finitary means. The idea behind reverse mathematics (and, presumably, the inspiration for its name) is to begin with a particular body of results believed to form the core of classical mathematics (and hence to be an indispensable part of the ideal systems of the Hilbertian), and thence to find the weakest possible set of axioms for proving these. In that way, one ends up with a set of axioms that is (at least conceptually speaking) equivalent to the theorems to be proven, rather than simply strong enough to entail them (and perhaps much stronger than what is really needed). One thus stands to eliminate the unnecessary strength present in the usual axiomatizations, and this may in turn enable him to more nearly approximate Hilbert's goal of a finitary soundness proof for ideal mathematics.

   (I am not so sure, however, that the notion of "necessary strength" tacitly employed by the advocates of reverse mathematics is the correct one. Simpson, at least, thinks that Hilbert was simply out to save classical mathematics. Hence, he takes the starting point of Hilbert's Program (and thus of reverse mathematics) to be some body of classical results, which must be preserved. I, however, regard this as incorrect. Hilbert did want to preserve classical mathematics, but this was not for him an end in itself. What he valued in classical mathematics was its efficiency (including its psychological naturalness) as a means of locating the truths of real or finitary mathematics. Hence, any alternative to classical mathematics having the same benefits of efficiency would presumably have been equally welcome to Hilbert. There may, of course, be no such alternatives, and it may even be that Hilbert believed this. But regardless of whether this is true (and who knows whether it is), it would still be misleading to describe Hilbert's goal as that of preserving classical mathematics.)

   Another strategy for getting a more accurate (and conservative) estimate of the Hilbertian's ideal commitments may be found in Detlefsen (1986: chs III, V). The idea there is also a sort of "reversing" strategy, one, however, which starts not from what is taken to be essential to classical mathematics (qua mathematics) but rather from what is judged to be the instrumentally useable portion of ideal mathematics. It begins with the belief that only some of the proofs in the usual ideal systems will be efficient enough to provide any advantage in efficiency to the Hilbertian instrumentalist; others will either be too long to be useable as an instrument, or they will not afford any advantage in efficiency over their real counterparts. In principle, such proofs could be eliminated from the ideal systems without any cost to the instrumentalist (i.e. without any loss in the efficiency of the ideal system). The fragment of the system remaining after such eliminations is what the instrumentalist is really responsible for, and his soundness proofs ought therefore ideally to be aimed at them rather than the usual systems. In essence, then, the idea is to start with the instrumentally gainful ideal proofs and work backward to a minimal system capturing them.

   Like the "reverse mathematics" of Friedman and Simpson, this approach to Hilbert's Program attempts to revive it by being more exact (and conservative) about the strength required for ideal mathematics. This concentration on the strength of the theories needed by the Hilbertian should be contrasted with our present emphasis, which is to raise the question of what happens to Hilbert's Program when one leaves the strength of the usual ideal theories intact, but alters the mode according to which they are generated. It may be, however, that the best way to develop Hilbert's Program would be to combine elements of both sorts of approaches; that is, to modify both the strength of the Hilbertian's ideal commitments and their mode of generation.

18 The locative and methodological elements may not, of course, be so cleanly defined. In particular, the locative element of a theory may also play a role in its methodological control over beliefs. Generally speaking, this will happen when there is a means of judging the quality of particular outputs of the locative element that is independent of that sponsored by the general description we have of the quality of its outputs. To take an example, we might believe of a set of wellconfirmed empirical generalizations that they are true and that the logic we use to derive observational consequences from them is truthpreserving. Still, we have an independent means (namely, observation) of evaluating those consequences. And if that independent means sponsors a contrary judgment, it might even cause us to reassess the judgment of quality based on the general qualitative description we have of the locative element (namely, that the generalizations concerned are true, and/or the logic for manipulating them truth-preserving). When part of our access to the epistemic quality of the locative element is of this "consequentialist" type, the general description we have of its epistemic quality will not form the whole of its methodological element.

   But though this is true and may even be typical of theories, it is important not to lose sight of the need for a description of the locative element (i.e. a methodology) that gives a general assessment of its output. For the whole idea behind a theory typically is that we either cannot or do not want to try to gain independent access (i.e. access not provided by following the prescribed method of belief-selection) to the outputs of the locative element. Thus, for example, an empirical theory is desirable just because we do not have timely and/or sufficiently safe and economical observational access to its observational consequences. Likewise, for the Hilbertian, ideal mathematics is desirable precisely because we do not have suitably efficient access to the truths of mathematics via

real proof. Thus, having independent access to the epistemic quality of the outputs of the locative element of a theory does not eliminate the need for a *general* assessment of its outputs.

19   Where does the proof referred to in (ii) come from? In particular, what sort of soundness principle does it come from? In the last section, we argued that the Hilbertian is not committed to the usual conservation condition since he is not responsible for proving of real sentences that are finitarily undecidable that they are provable by ideal means only if they are also provable by real (i.e. finitary) means. Thus, he need not show of each real sentence formulable in the language of a given ideal theory T that it is provable in T only if it is also provable by real means. This restricted soundness principle, however, can only give rise to a proof of the sort referred to in (ii) when it can also be shown that the real sentence for which one has an ideal proof (by (i)) is decidable by real means, and this is something the ideal theorizer would like to avoid having to show (since he has no general finitary procedure for doing so).

There is, however, a more congenial alternative modification of the usual conservation condition which would seem to be available. The idea behind this alternative is that the Hilbertian has no need to show of real sentences that are *not* decidable by the *ideal* theory he is defending that they are provable in it only if they are also provable by real means. This is so because he will most assuredly not make use of any ideal proofs of such sentences; and if he cannot make use of them, there is no reason why they should be included within the scope of his soundness condition. Modified accordingly, the Hilbertian's obligation reduces to that of showing (finitarily), of each real sentence decidable in the ideal theory he is defending, that it is provable in that theory only if it is provable by real means. Such a restricted soundness condition, in concert with proof referred to in clause (i), affords the Hilbertian a proof of the sort referred to in (ii).

20   For a refinement and extension of these brief remarks concerning the threat of epistemic "dilution" and its place in the Hilbertian's thinking, see Detlefsen (1986: chs I, II).

21   Note that here R need not be identified with finitary reasoning *per se.* It is only that body of finitary reasoning that is to be replaced by the ideal system T, and this might not constitute the whole of finitary reasoning. Indeed, the Hilbertian is not committed to holding that the whole of finitary reasoning ought to be replaced; there might, after all, be cases of finitary reasoning whose efficiency cannot be improved upon by any known ideal means. These possible differences between R and the whole body of finitary reasoning also indicate that it would be a mistake to demand (by way of a soundness principle) of the real sentences provable in T that they should be provable in R.

22   To show this, one appeals to (a) G2 for T, (b) a premise to the effect that if the theorem-wise coextensivity of T and T′ were provable in T, then "Con(T′) → Con(T)" would also be provable in T, and (c) the fact that Con(T′) is provable in T.

23   Note that even with standard (as opposed to consistency-minded) theories, one does not have finitary control over the locative element, since one cannot show finitarily that they do not prove some absurdity.

24   Hilbert's Program, as I understand it, requires only a proof of the realsoundness of the *instrumentally gainful* ideal methods (i.e. those ideal derivations which are (i) short and simple enough to be of some use by agents with our cognitive limitations (possibly with the assistance of realizable computing machines), and (ii) more efficient than any available real proof of the same result). This, however, requires that we have some sort of "complexity metric" for rating (and comparing) the complexity of real and ideal proofs. There are, of course, various complexity metrics that have found their way into the proof-theoretic literature, and the recent literature in theoretical computer science has produced even more. Yet all these complexity metrics seem to be designed to measure a general type of complexity that might be called "verificational complexity"; that is, the type of complexity that is encountered in determining of a given syntactical entity whether or not it is a proof in a given system of proofs. It seems, however, that what the Hilbertian is chiefly concerned with is not verificational, but rather "inventional" complexity; that is, the type of complexity that is encountered in coming up with a proof in the first place (as opposed to verifying of a *given* item that it is a proof). This is strongly suggested by Hilbert's statement (1927:475) that his

formula game is carried out according to definite rules, in which the *technique of our thinking* is expressed. These rules form a closed system that can be discovered and definitively stated. The fundamental idea of my proof theory is none other than to describe the activity of understanding, to make a protocol of the rules according to which our thinking actually proceeds. Thinking, it so happens, parallels speaking and writing: we form statements and place them one behind another. If any totality of observations deserves to be made the object of a serious and thorough investigation, it is this one.

If this way of looking at Hilbert's Program is right, then, in order to properly evaluate it, a metric for *inventional* complexity would have to be developed. Since, however, we are far from being able to do this, we are equally far from being able to give a definitive evaluation of Hilbert's Program—the traditional negative arguments and more recent positive proposals (e.g. reverse mathematics) notwithstanding. Much basic philosophical work remains to be done.

REFERENCES

Arai, T. (1990) "Derivability conditions on Rosser's provability predicates," *Notre Dame Journal of Formal Logic* 31:487–97.
Detlefsen, M. (1986) *Hilbert's Program,* Dordrecht: Reidel.
Feferman, S. (1960) "The arithmetization of metamathematics in a general setting," *Fundamenta Mathematicae* 49:35–92.

Gentzen, G. (1936) "The consistency of elementary number theory," reprinted in M.E.Szabo (trans, and ed.) *The Collected Works of Gerhard Gentzen,* Amsterdam: North-Holland, 1969.

——(1938) "The present state of research into the foundations of mathematics," reprinted in M.E.Szabo (trans, and ed.) *The Collected Works of Gerhard Gentzen,* Amsterdam: North-Holland, 1969.

Guaspari, D. and Solovay, R. (1979) "Rosser sentences," *Annals of Mathematical Logic* 16:81–99.

Hilbert, D. (1901) "Mathematical problems," *Bulletin of the American Mathematical Society* 8:437–79.

——(1925) "On the infinite," reprinted in J.van Heijenoort (ed.) *From Frege to Gödel,* Cambridge, MA: Harvard University Press, 1967.

——(1927) "The foundations of mathematics," reprinted in J.van Heijenoort (ed.) *From Frege to Gödel,* Cambridge, MA: Harvard University Press, 1967.

——(1930) *Grundlagen der Geometrie,* 7th edn, Leipzig and Berlin: Teubner.

——(1931) "Die Grundlegung der elementaren Zahlenlehre," in *Gesammelte Abhandlungen,* vol. 3, Berlin: Julius Springer Verlag, 1935.

Jeroslow, R. (1975) "Experimental logics and Π^0_2 theories," *Journal of Philosophical Logic* 4:253–67.

Kreisel, G. (1971) "A survey of proof theory II," in J.E.Fenstad (ed.) *Proceedings of the Second Scandinavian Logic Symposium,* Amsterdam, North-Holland.

——(1976) "What have we learnt from Hilbert's Second Problem?," *A MS Proceedings of Symposia in Pure Mathematics,* vol. 28, Providence, RI: American Mathematical Society.

——(1980) "Kurt Gödel," *Biographical Memoirs of the Fellows of the Royal Society* 26:149–224.

Kreisel, G. and Takuti, G. (1974) "Formally self-referential propositions for cut-free analysis and related systems," *Dissertationes Mathematicae* 118:4–50.

Prawitz, D. (1981) "Philosophical aspects of proof theory," in *Contemporary Philosophy: A New Survey,* vol. 1, The Hague: Martinus Nijhof.

Rosser, J.B. (1936) "Extensions of some theorems of Gödel and Church," *Journal of Symbolic Logic* 1:87–91.

Sieg, W. (1985) "Fragments of arithmetic," *Annals of Pure and Applied Logic* 28:33–71.

Simpson, S. (1988) "Partial realizations of Hilbert's Program," *Journal of Symbolic Logic* 53:349–63.

Smorynski, C. (1977) "The incompleteness theorems," in J.Barwise (ed.) *Handbook of Mathematical Logic,* Amsterdam: North-Holland.

——(1985) *Self-Reference and Modal Logic,* New York: SpringerVerlag.

——(1988) "Hilbert's Programme," Logic Group Preprint Series No. 31, Department of Philosophy, University of Utrecht (to appear as the introduction to the fourth chapter of a book in progress on the metamathematics of arithmetic).

Visser, A. (1989) "Peano's smart children: a provability logical study of systems with built-in consistency," *Notre Dame Journal of Formal Logic* 30:161–96.

INDEX