

---

# Departing from Frege

---

Essays in the philosophy of language

R. M. Sainsbury



London and New York

**Also available as a printed book  
see title verso for ISBN details**

---

# Departing from Frege

---

Gottlob Frege is now regarded as one of the world's greatest philosophers, and the founder of modern logic. In addition to his work on the foundations of mathematics, his writing on sense and reference remains deeply influential.

*Departing From Frege* takes Frege's work as a point of departure, but argues that we must depart considerably from Frege's own views if we are to work towards an adequate conception of natural language. Mark Sainsbury suggests that the two aspects that are most important to retain in Frege's work are the distinction between sense and reference, and the possibility of sense without a referent. Many philosophers today take the equation of sense with "mode of presentation of an object" or "way of thinking of an object" as definitive of Fregeanism. Although some of Frege's words do suggest this view, Sainsbury argues that it is not required for the distinction between sense and reference. Moreover, it makes it difficult to accommodate sense without a referent, and impossible to do justice to Frege's clear commitment to sense as the common property of all who speak a language.

In this selection of essays, Mark Sainsbury brings a new position into view. It shares with current "direct reference" theories the rejection of descriptivist accounts, but differs from them in its requirement that some expressions with the same referent be given different semantic descriptions, and it allows for sense without a referent. *Departing From Frege* is an outstanding contribution to philosophy of language and logic and will be invaluable to all those interested in Frege and the philosophy of language.

**Mark Sainsbury** is Susan Stebbing Professor of Philosophy at King's College London. He is author of *Russell* (Routledge, 1985), *Paradoxes* (1995) and *Logical Forms* (2000). He is also a former editor of *Mind*.

---

## International Library of Philosophy

Edited by José Luis Bermúdez, Tim Crane and Peter Sullivan

*Advisory Board: Jonathan Barnes, Fred Dretske, Frances Kamm,  
Brian Leiter, Huw Price and Sydney Shoemaker*

---

Recent titles in the ILP:

**The Facts of Causation**

*D. H. Mellor*

**The Conceptual Roots of Mathematics**

*J. R. Lucas*

**Stream of Consciousness**

*Barry Dainton*

**Knowledge and Reference in Empirical Science**

*Jody Azzouni*

**Reason Without Freedom**

*David Owens*

**The Price of Doubt**

*N. M. L. Nathan*

**Matters of Mind**

*Scott Sturgeon*

**Logic, Form and Grammar**

*Peter Long*

**The Metaphysicians of Meaning**

*Gideon Makin*

**Logical Investigations, Vols I & II**

*Edmund Husserl*

**Truth Without Objectivity**

*Max Kölbel*

**Departing from Frege**

*R. M. Sainsbury*

---

# Departing from Frege

---

Essays in the philosophy of language

R. M. Sainsbury



London and New York

First published 2002

by Routledge

11 New Fetter Lane, London EC4P 4EE

Simultaneously published in the USA and Canada

by Routledge

29 West 35th Street, New York, NY 10001

This edition published in the Taylor & Francis e-Library, 2003.

© 2002 R. M. Sainsbury

All rights reserved. No part of this book may be reprinted or reproduced or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

*British Library Cataloguing in Publication Data*

A catalogue record for this book is available from the British Library

*Library of Congress Cataloging in Publication Data*

A catalog record for this book has been requested

ISBN 0-203-21893-0 Master e-book ISBN

ISBN 0-203-27419-9 (Adobe eReader Format)

ISBN 0-415-27255-6 (Print Edition)

---

# Contents

---

<i>Acknowledgements</i>	ix
<b>Introduction: Departing from Frege</b>	<b>1</b>
1 Paring down Frege	1
1.1 <i>Modes of presentation</i>	2
1.2 <i>Specifying senses</i>	4
1.3 <i>Senses and indirect speech</i>	7
1.4 <i>Indexicality</i>	9
1.5 <i>Sense without a referent</i>	9
1.6 <i>The determination of reference, sharp boundaries, saturation</i>	14
1.7 <i>The nature of sense</i>	17
1.8 <i>An overlooked position</i>	19
2 The essays: provenance and comments	<b>19</b>
<b>I Understanding and theories of meaning (1979)</b>	<b>29</b>
1 Understanding and structure	30
2 Extensionality	35
3 “Means that” and theories of meaning	37
Postscript: on truth and “truth conditions”	41
<b>II Evans on reference (1985)</b>	<b>43</b>
1 Singular terms	43
2 Information, communication, thought	48
2.1 <i>Information-invoking singular terms and information-based thought</i>	48
2.2 <i>Communication</i>	55
2.3 <i>Two objections to Evans’s view of thoughts</i>	57
3 Existence	64
4 Coda	69

<b>III</b>	<b>Concepts without boundaries (1990)</b>	<b>71</b>
1	Classification by classes	71
2	There is no set of red things	72
3	An unsharp boundary is many sharp boundaries	73
4	Fuzzy logic and supervaluations	75
5	Boundaryless concepts	76
6	Paradox	78
7	Semantics for boundarylessness	79
8	Paradigms and prototypes	81
9	Conclusion	83
<b>IV</b>	<b>Russell on names and communication (1993)</b>	<b>85</b>
1	Introduction	85
2	What Russell actually said	87
3	Russell's theory of communication and the public reference of names	90
4	Russell's theory of the truth conditions of utterances containing names	93
	<i>Comments</i>	93
5	Solutions to some problems	95
	<i>Existence</i>	95
	<i>Belief</i>	97
	<i>Identity</i>	97
6	Criticisms of description theories	98
<b>V</b>	<b>How can some thing say something? (1996)</b>	<b>102</b>
1	Background	102
2	How to solve the problem: Candlish's suggestions	104
3	Another approach	106
<b>VI</b>	<b>Easy possibilities (1997)</b>	<b>112</b>
1	Introduction	112
2	Tracking	113
2	KK	115
4	Hidden snags	116
5	Determinism	117
6	Worlds	118
7	Vagueness	120
8	Objections	122
<b>VII</b>	<b>Fregean sense (1997)</b>	<b>125</b>
1	Introduction	125
	<i>1.1 Sense is a property of tokens</i>	125
	<i>1.2 Sense without senses</i>	126

1.3	<i>Sense and explanation</i>	126
2	Millikan's objections to Fregean sense	127
2.1	<i>Meaning rationalism</i>	127
2.2	<i>Identity and folders</i>	131
2.3	<i>Identity, recognition and the apriori</i>	132
2.4	<i>Psychology and semantics</i>	133
3	Logic and guaranteed sameness of reference	133
<b>VIII</b>	<b>Indexicals and reported speech (1998)</b>	<b>137</b>
1	A hypothesis about meaning	137
2	Non-detachability and semantic theory	140
3	Indexicality, anaphora and reducibility	146
4	Reducibility: action, science, time and God	149
	<i>Reducibility</i>	149
	<i>Action</i>	150
	<i>Science</i>	152
	<i>Time</i>	152
	<i>God</i>	154
5	The Fregean connection	155
<b>IX</b>	<b>Names, fictional names, and "really" (1999)</b>	<b>159</b>
1	Introduction	159
2	Evans's arguments for the Russellian view of names	159
2.1	<i>Discriminating knowledge</i>	160
2.2	<i>Invoking information</i>	161
2.3	<i>Understanding and knowledge</i>	162
2.4	<i>The first argument: truth is seamless</i>	162
2.5	<i>The second argument: the demand of unity</i>	164
3	Understanding fiction	167
3.1	<i>Quasi-fear</i>	167
3.2	<i>Make-believe and understanding</i>	169
4	"Really"	171
<b>X</b>	<b>Knowing meanings and knowing entities (2001)</b>	<b>181</b>
1	Acquaintance with meanings	181
2	$x$ means $y$	184
3	"That"-clauses do not refer to meanings	185
4	Meanings and understanding	186
5	Indirect questions and knowing what	188
6	Things which are not entities?	190
<b>XI</b>	<b>Two ways to smoke a cigarette (2001)</b>	<b>192</b>
1	Introduction	192



2	Understanding utterances: covert indexicality as a resource for compositionalist	193
3	Apparent counterexamples to compositionality	195
4	Applying the strategies	197
4.1	<i>Adjectival modification</i>	197
4.2	<i>Genitives</i>	199
4.3	<i>Compounded nouns</i>	200
4.4	<i>Verbs</i>	201
4.5	<i>The Travis effect</i>	201
5	Conclusion	203
<b>XII</b>	<b>Sense without reference (2001)</b>	<b>205</b>
1	Introduction	205
2	Some general Fregean considerations	208
3	Name-using practices	210
3.1	<i>A simple model</i>	211
3.2	<i>Duplicated sources: qualitatively the same information may inform different practices</i>	212
3.3	<i>Information involved in distinguishing practices?</i>	213
3.4	<i>Fusion</i>	215
3.5	<i>Confusions</i>	219
3.6	<i>Fission</i>	220
4	Source a disjunctive property?	222
	<i>Bibliography</i>	224
	<i>Index</i>	231

---

# Acknowledgements

---

The essays are reproduced with kind permission from the following sources.

I “Understanding and theories of meaning”. *Proceedings of the Aristotelian Society* **80**, 1979: 127–44. Presented at a meeting of the Aristotelian Society in Tavistock Place, London WC1, on Monday 3 March, 1980, at 6:30 p.m. Reprinted by courtesy of the Editor of the Aristotelian Society: © 1979.

II “Evans on reference”. Originally published as R. M. Sainsbury, “*Critical Notice: The Varieties of Reference* by Gareth Evans”, *Mind* **94**, 1985: 120–42. Reprinted by kind permission of the Mind Association.

III “Concepts without boundaries”. London: King’s College London, 1990. This is the text of the philosophical part of my inaugural lecture as Susan Stebbing Professor of Philosophy at King’s College London, delivered on 6 November 1990; hence the absence of references and footnotes. It was reprinted in Rosanna Keefe and Peter Smith (eds) *Vagueness: A Reader*, Cambridge MA: MIT Press, 1996: 251–64.

IV “Russell on names and communication”. In Andrew Irvine and Gary Wedeking (eds) *Russell and Analytic Philosophy*, Toronto: University of Toronto Press, 1993: 3–21. Reprinted with permission of University of Toronto Press. The volume collects papers presented at a conference on Russell held at the University of British Columbia, Vancouver, in 1992.

V “How can some thing say something?”. In Ray Monk and Anthony Palmer (eds) *Bertrand Russell and the Origins of Analytic Philosophy*, Bristol: Thoemmes, 1996: 137–53. Reprinted by permission of Thoemmes Press. The volume collects papers from a conference on Russell held at the University of Southampton in July 1995, which included a symposium between Stuart Candlish and myself. Without informing me, the editors published my paper under a title of their choosing (“How can we mean something?”).

VI “Easy possibilities”. *Philosophy and Phenomenological Research* **57**, 1997: 907–19. Copyright © International Phenomenology Society 1997. Reprinted with permission.

VII “Fregean sense”. In Timothy Childers, Petr Kolář and Vladimir Svoboda (eds) *Logica '96*, Prague: Filosofia, 1997: 261–76. The volume collects papers presented at the annual Logica conference held at Liblice Castle in the Czech Republic, June 1996.

VIII “Indexicals and reported speech”. *Philosophical Logic: Proceedings of the British Academy* **95**, 1998: 45–69. Edited by Timothy Smiley. Reprinted by permission of Oxford University Press. The volume collects papers presented at a British Academy conference on philosophical logic on 16 March 1997, organized by Professor Timothy Smiley. Jimmy Altham was the co-symposiast.

IX “Names, fictional names and ‘really’”. *Proceedings of the Aristotelian Society, Supplementary Volume* **73**, 1999: 243–69. Reprinted by courtesy of the Editor of the Aristotelian Society: © 1999. This comes from a symposium at the July 1999 Joint Session of the Mind Association and the Aristotelian Society. David Wiggins was my co-symposiast: hence the invitation at the end of my paper for him to expand his views.

X “Knowing meanings and knowing entities”. First published in Uxe Meixner and Peter Simons (eds) *Metaphysics in the Post-Metaphysical Age: Proceedings of the 22nd International Wittgenstein Symposium, 1999*, Vienna, 2001: 106–15. Copyright © öbvahpt Verlagsgesellschaft, Vienna, 2001. Reprinted with permission.

XI “Two ways to smoke a cigarette”. *Ratio* **14**, 2001: 386–406. Reprinted with the permission of Blackwell Publishing. The original version of this paper was published alongside two papers on compositionality, one by Paul Horwich (2001) and one jointly by Jerry Fodor and Ernest Lepore (2001), and contained a discussion of their disagreement. The present reprinting excises that material and makes some consequential adjustments.

XII “Sense without reference”. In A. Newen, U. Nortmann and R. Stuhlmann-Laeisz (eds), *Building on Frege*, 2001: 211–30. Stanford, CA: CSLI Publications, 2001. Copyright © 2001 by CSLI Publications. Used with permission of CSLI Publications. The volume collects papers delivered at a conference on Frege in Bonn in October 1998.

# Departing from Frege

---

For most philosophers who write about natural language and meaning, Frege's work has served as a point of departure: a place to start, and an origin from which to measure divergence along a variety of axes. At least until his very last work, Frege was officially not concerned with the semantics of natural languages, and the majority of his writing relates to the foundations of mathematics. Even so, much of the theorizing about meaning in the last fifty or more years could be described as a series of responses to Frege's views about sense and reference. Probably nobody would now defend precisely the position Frege developed, but many locate their own views in terms of the respects in which, and the motivations for which, they depart from Frege. Even those who take themselves to be wholly opposed to Frege regard it as compulsory to address the data, or supposed data, that Frege adduced in support of his distinction between sense and reference and related doctrines. At the centre of concern have been singular terms, and in particular the question of how distinct coreferring ones can, apparently, behave differently, and how they behave in ascriptions of propositional attitude.

The essays reprinted here fit this general picture. Frege is taken as the point of departure, and many of the essays, implicitly or explicitly, explore what can be rejected while retaining what I take to be the crucial elements: that some coreferring singular terms make different semantic contributions, and so require different semantic accounts; and that a singular term may be perfectly intelligible yet lack a referent. In the first part of this introduction I develop the general theme of paring down Frege. In the second part I comment on the individual essays.

### **I Paring down Frege**

Frege's doctrine of sense and reference starts from the idea that some public and shared differences among things known cannot be attributed to the referents of the words involved in the expression of the knowledge. Though this

starting point has been challenged, I take it as given.<sup>1</sup> It is one thing to know that Hesperus is Hesperus and another to know that it is Phosphorus, one thing to know that Hesperus is visible, another to know that Phosphorus is. These are different items of knowledge, each item is potentially accessible to anyone, and so is the fact that the items are distinct. This I shall call “Frege’s datum”. When we move from things known to the words that express them, it seems that an adequate semantic description should capture a difference between words which, though having the same referent, can contribute differently to the expression of knowledge. The adequacy condition for a semantic description of words is not merely that they be assigned their actual referent. If this were correct, it would be adequate to use just the same semantic description of both “Hesperus” and “Phosphorus”, in which case the semantic description would miss whatever the difference is which enables them to make different contributions to the expression of knowledge. The needed difference in an adequate semantic description marks a difference of sense. This line of thought is what I regard as one constituent of the core of Fregeanism; the other is his recognition of the possibility of sense without reference. Much of the material which surrounds these core constituents must, I think, be set aside, including some aspects, like modes of presentation, which have often been taken to be inseparable from the core.

### ***1.1 Modes of presentation***

Frege initially characterized the sense of a singular term as the mode of presentation of its referent, and introduced the notion of a mode of presentation by a more or less perceptual example: a point where lines *A*, *B* and *C* intersect may be presented as the point of intersection of *A* and *B* or as the point of intersection of *B* and *C*. Similarly, the planet Venus may be presented in different ways, as the first heavenly body to appear at dusk or as the last to fade at dawn. Whatever value modes of presentation may have as a quick intuitive indicator of what sense is supposed to be, they are damaging in many ways, and are to be dropped from the kind of pared down Fregeanism I wish to support (on modes of presentation, see especially essays VII: “Fregean sense” and VIII: “Indexicals and reported speech”). The following are among the reasons for this rejection:

- 1 If sense is the mode of presentation of an object, it is hard to see how sense without reference is possible.
- 2 Modes of presentation are supposed to be typically available independently of language, so that one would expect that the sense of an

<sup>1</sup> Sainsbury (1983) indicates briefly how it might be challenged. The challenge is given theoretical depth by Salmon (1986).

expression can typically be stated without making use of that expression. This imports a reductionist dynamic into many Fregean and post-Fregean positions, which I regard as unwelcome.

- 3 As Frege's doctrine developed, senses became abstract inhabitants of the third world (the world composed neither of things, perceivable by the senses, nor of ideas, which require an owner in order to exist: Frege 1918: 360–2). In that case, a story needs to be told about how we recognize them, and how we know how many senses are involved when we engage in more than one act of reference in thought. If Frege's datum is not to arise again for senses, they need to be objects concerning which we have infallible recognitional powers, and no one nowadays would take such objects seriously. Moreover, if senses are entities, understanding is naturally construed as knowledge of entities, a view that I argue against in essay X ("Knowing meanings and knowing entities").
- 4 As Frege was aware, his view of sense as mode of presentation made it hard to find public senses, and so hard to regard sense as something which should feature in a semantic description of expressions in a public language.

The last difficulty is familiar, and Frege in effect drew attention to it in the famous footnote about "Aristotle" in "On sense and reference" (1892a). By the time of "Thoughts" (1918), his position on this matter was close to a *reductio ad absurdum*. He writes:

with a proper name, it is a matter of the way that the object so designated is presented. This may happen in different ways, and *to every such way there corresponds a special sense* of a sentence containing the proper name. So we must really stipulate that for every proper name there shall be just one associated manner of presentation of the object so designated. (Frege 1918: 359, my emphasis)

In the first two sentences of the quotation, Frege recognizes that the doctrine of sense as mode of presentation leads to the unpalatable conclusion that each proper name, even in its use for a single referent, has no public sense, in contradiction to his declared opinion that "the sense of a proper name is grasped by everybody who is sufficiently familiar with the language" (Frege 1892a: 158). The second part of the quotation constitutes a hasty attempt to modify the original theory: though each proper name is associated with many modes of presentation, it is a semantic stipulation that only one such mode constitutes its sense. Completing the theory would require that for each name, some stipulation singles out the favoured mode of presentation. But Frege himself evinces no confidence that this is so, and Saul Kripke's examples have convinced almost everyone that it is not (Kripke 1972).

## 1.2 Specifying senses

So long as modes of presentation are offered as what constitutes senses, it is natural to suppose that specifying a sense would involve some revealing specification of the relevant mode of presentation. Frege himself was ambivalent. As discussed, the mode of presentation story sometimes led him to regard specifying a sense as making a substantive claim about modes of presentation. But one pronouncement is more modest:

In order to speak of the sense of an expression “A” one may simply use the phrase “the sense of the expression ‘A’”. (Frege 1892a: 159)

Here Frege commits firmly to a unique sense, and eschews any attempt at a revealing specification of a mode of presentation. This seems to me the right approach, though this is not everyone’s view about what makes for a “Fregean” specification of sense. For example, Michael Dummett has suggested that sense would constitute an epistemic route to the referent; and many philosophers have supposed that the Fregean sense of a proper name can be given by a definite description.

Frege and Russell have been lumped together as “description theorists” with respect to what we ordinarily count as proper names. The attribution to Frege is based largely on the footnote about Aristotle in “On sense and reference”, and that to Russell on too hasty a reading of various remarks about names being abbreviated or truncated descriptions. Russell’s position stands in interesting contrast to Frege’s: for Russell, definite descriptions do indeed have a part to play in characterizing the use of proper names (ordinarily so-called), for we must use a definite description if we are to give explicit expression to

the thought in the mind of a person using a proper name correctly . . . Moreover the description required to express the thought will vary for different people, or for the same person at different times. (Russell 1912: 29)

The variability shows that Russell did not suppose that the definite descriptions could be used to specify a constant public sense. This opens up for Russell the possibility of a two-level account, which I describe in essay IV (“Russell on names and communication”): at one level there is thought, which is largely descriptive; at another, there is what is communicated, which may feature an object, even though it is not one which we can name (because we cannot be acquainted with it). As Russell put it: “we can . . . describe the proposition we should like to affirm”, which is one of which an unknown object is a constituent (that is, an object with which we are not acquainted). Russell thus feels under no compulsion to move from a descriptive account of thought to a descriptive account of the unit of communication.

Something very like Russell's position is argued for by Evans (1982), as discussed in more detail in essay II below. Evans accused Frege of running together two ideas: that of what is conventionally communicated by the utterance of a sentence, and that of what is going on in the mind of the utterer. Any reasonable view must allow that a full description of an utterer's mental state will contain more than just the public content of the utterance; Frege's "associated idea" marks the position for such additional material within his theory. However, Russell's position is instructive, as it represents one internally consistent way of incorporating something like modes of presentation. Russell's idiosyncratic and changing associated descriptions could well be described as modes of presentation or ways of thinking of their object, but plainly do not correspond to anything public and shared between speakers. The object itself is shared, but is not appropriate to an account of what is happening in the mind of the speaker. There is no room within Frege's system for such a split between the unit of thought and the unit of communication, though Frege struggles with the resulting problems in his discussion of "I"-thoughts, comparing the subject's point of view with the interpreter's. If modes of presentation are accorded the central place which most Fregeans accord them, then there is a natural tendency to see Frege as committed at least to something similar to a descriptivist view of thought. If the mode of presentation is implicit in the proper name, then presumably it should be theoretically possible to make it explicit, and it is hard to imagine how this could be done without using a definite description. Even if Dummett is right to say that "there is nothing in what he [Frege] says to warrant the conclusion that the sense of a proper name is always the sense of some complex description" (Dummett 1973: 97–8), it remains that a Fregean who takes modes of presentation as central will find it hard to avoid commitment to the descriptivist conclusion which Dummett disavows.

I propose that we set modes of presentation aside, and use Frege's modest suggestion recently quoted: we can specify the sense of an expression, "A", by using the phrase "the sense of expression 'A'". A general point in favour of this more modest approach to sense specification is that we would expect a typical word in a language to differ in sense not only from other words, but also from other phrases. At any event, we could certainly imagine a language which totally lacked this kind of redundancy. In such a redundancy-free language, no word would have the same sense as any other word or phrase. A semantic theory for this language couched in the language itself, and which aimed to specify the sense of each word in the language, could but reuse the target word in the specification of that word's sense. The semantic theory would gravitate towards "homophonic" specifications. Logicians are used to this in model-theoretic clauses for logical constants; it should be regarded as normal across the board, departures being justified by special features of language, like indexicality (see §1.4 below, and essay VIII: "Indexicals and reported speech").



Such an approach has been championed by John McDowell, using Donald Davidson's truth theoretic framework. Davidson suggested that a semantic description of a language could take the form of a truth-theory selected for its potential to contribute to the task of radical interpretation. The strategy is to shed light on the nature of meaning in general by considering how a formalized theory (a set of sentences closed under deduction), fit to somehow state or fix the meanings of the sentences of some specific language, can be constructed and empirically justified. In the terminology of essay I ("Understanding and theories of meaning"), the strategy is to use general reflections on the nature of theories<sub>1</sub> of meaning in the service of building a theory<sub>2</sub> of meaning.

A theory<sub>1</sub> of meaning will not say anything of interest about the meaning of words, and will make no contribution to their analysis; homophony will be the rule. An axiom for "green" will be based on something like: "'green' is true just of green things", and one for "Hesperus" on something like "'Hesperus' stands for Hesperus". One merit of Davidson's approach is that it refocuses attention from the unprofitable attempt to "give the meaning of a word" to two other matters with which one can reasonably hope for more progress: getting a theoretical grip on the combinatory devices of language, those which give rise to its compositionality; and exploring the way in which the norms of linguistic interpretation connect with facts about what its speakers use language to do.

The distinction between sense and reference should give rise to no expectation that homophony will be other than the norm; the general considerations recently mentioned, concerning how one might hope to specify the sense of words in a language free of redundancy, still apply: McDowell (1977) argues on Frege's behalf that it is quite in keeping with the distinction between sense and reference to specify the sense of an expression by reusing it in an axiom of the kind just described: "Hesperus" stands for Hesperus. Natural reactions are that, since this merely specifies a referent, it cannot in addition specify a sense; and that since Hesperus is Phosphorus, the approach will ascribe to "Hesperus" just what it ascribes to "Phosphorus", and so will miss the sense/reference distinction. McDowell has shown that these responses are misguided. An axiom of the envisaged homophonic kind will lead to theorems which can be used in good interpretations of speech. But a theory with the axiom "'Hesperus' stands for Phosphorus" will lead to bad interpretations. For example, it will lead to an interpretation of an utterance like "Hesperus is visible but Phosphorus is not" as a manifest contradiction, which it does not seem to be. The sense/reference distinction is marked within the Davidsonian framework by the selection of the right way to specify the referent: a correct axiom will state the reference, and thereby show the sense.

McDowell's point that a Fregean can handle sense and difference of sense without departing from homophonic specifications of sense is crucial to the success of the pared down Fregeanism which I envisage. However, in the

same article, McDowell also suggests that these senses are object-involving or *de re*: they essentially require a referent in order to be intelligible. This sets aside Frege's favourable opinion about the possibility of sense without reference, is logically quite independent of McDowell's arguments for homophonic specifications, and leads to a semantics that is at variance with the facts. These issues are taken up in §1.5 below and in essay XII ("Sense without reference").

### 1.3 Senses and indirect speech

Frege needs senses as entities for his account of the semantics of propositional attitudes. He held that when, for example, we ascribe a belief to someone, we should aim to specify the thought which that person entertained. This led him to the view that the referent of the sentence following "that" in such an ascription is a thought; that is, in this context the referent of the sentence is what is customarily its sense. It also seemed to him that the supposition that, in general, expressions in these contexts refer to their customary senses would correctly predict truth preserving substitution conditions: the fact that, in these contexts, sameness of customary sense is a minimally sufficient condition for expressions to be exchanged without change of truth value would be simply a special case of the insensitivity of the referent of a whole to replacement of a component expression by another with (in the context) the same referent.

Frege himself held to the distinction between sense and reference even for indirect contexts, like those which occur in ascriptions of propositional attitudes. Dummett has argued that this leads to an indefinite hierarchy of senses, indirect senses, doubly indirect senses and so on, as sentences are embedded more and more deeply in the iterable idioms of indirect speech ("Jack believes that Jill believes that Jack believes that . . . etc"). He suggests a simplification which every Fregean should be pleased to adopt: the distinction between sense and reference lapses in indirect contexts, so that indirect sense equals indirect referent equals customary sense. This is a step forward, but it leaves a position which is still reliant upon senses as entities.

If senses pulled their weight as entities in an account of indirect discourse, we would have a reason to accept them as entities. But in fact they do not. Indeed, even the same sense relation does not do what it ought to do within the Fregean system, as Benson Mates (1950) has shown. Sameness of customary sense does not guarantee substitutivity *salva veritate* in every context. It cannot be ruled out a priori that a language should contain pairs of expressions with the same customary sense. Suppose that "Greek" and "Hellene" are such a pair. We may be confident that every pair of sentences like "Jack believes that Yannis is a Greek" and "Jack believes that Yannis is a Hellene" must have the same truth value; yet we must recognize that some people, perhaps philosophers in particular, are capable of doubting even what strike

most people as plain truths. These philosophers may hold that whereas there is no room to doubt whether everyone believes that all Greeks are Greeks, there is room to doubt whether everyone believes that all Greeks are Hellenes. To the extent that the story about these (hypothetical) philosophers is not manifestly contradictory, we have a demonstration that the substitution of expressions alike in sense does not always preserve truth value. The same sense relation cannot do all that is needed of an account of the semantics of attributions of propositional attitude. It cannot be denied that this detracts from the power and sweep of Frege's system.

The adoption of a Davidson-style paratactic account of the logical form of ascriptions of propositional attitudes is consistent with Frege's main tenets. In particular, it is consistent with the distinction between sense and reference, and with the possibility of sense without reference. When it comes to analysis of the first sentence in these paratactic logical forms, for example "Galileo said that.", Davidson invokes the samesaying relation. In essay VIII ("Indexicals and reported speech") I consider the idea that we can get a fix on sameness of sense in terms of the criteria we adopt for counting reports of speech as correct or incorrect: in short, in terms of samesaying. This is typical of the inversion of traditional Fregean priorities which I think is appropriate. Whereas, traditionally, the Fregean would "account for" or "explain" the correctness of reports of speech in terms of the same sense relation, I think that this places too much confidence in the antecedent availability of the notion of sense, and that a more modest aim is more appropriate because attainable: to take as data our judgements about the correctness of reports of speech, and use these to animate the conception of sense. The purported traditional explanation is useless, for two reasons: first, as we have seen, it leads to incorrect predictions in Mates-type cases; and secondly the notion of sense can neither be taken for granted as a primitive, nor defined by any single phenomenon. It is therefore best to see it as informed by a variety of matters, including how we count items of knowledge, what we regard as *manifestly* contradictory, our justification for treating equiform tokens as alike in reference (see essay VII "Fregean sense"), and our criteria for judging the correctness of reports of speech.

For Frege, and many subsequent writers, idioms of propositional attitude form a homogeneous class. On the present view this is not so. Whereas there is some plausibility in allowing judgements of sameness of sense to be guided by correctness of speech reports, there would be no plausibility in using correctness of belief reports as such a guide: these reports are too much at the mercy of a variety of pragmatic factors, and would give conflicting guidance depending upon contextually determined considerations. Moreover, there may not even be syntactic uniformity in the class (see essay X, note 5).

### 1.4 Indexicality

Frege's distinction between sense and reference does not apply straightforwardly to indexical expressions. Expressions with the same sense have the same referent. But different utterances of the same indexical expression, "that" for example, have different referents. Hence the expressions themselves, that is the expression types, do not have a Fregean sense.

The "semantic rule" with which an indexical is associated, its "character" in Kaplan's sense, has to be regarded not as its Fregean sense, but as some sort of rule for determining the sense of an arbitrary use of it. This means that, where the difference matters, the sense–reference distinction belongs to expression tokens, not types. This has some far-reaching consequences, some of which are spelled out in essay VIII. One important issue is whether there are "indexical thoughts" in the sense of thoughts which, for some indexical expression, can only be specified by employing that expression. If "I"-thoughts are indexical in this sense, they are inaccessible to all save the thinker; and Frege has been read as committed to just this view. An alternative is that there are indexical thoughts only in the weaker sense of thoughts which, at the point at which they are available for guiding an action, will typically be expressed by the agent by means of an indexical. This is consistent with the very same thought being expressible without the use of an indexical, a thesis argued for in essay VIII. It may well be that the expression of such thoughts by those other than the agent must exploit anaphoric dependence upon something outside the content-giving part of the attribution. Anaphoric dependence and indexicality are related phenomena. The critical difference is that the former poses no threat to the public character of thought. Even if only Jack can use "I" to think the thought that he expresses by "I am now late", others can, I argue, think the same thought using other words: at such-and-such a time, Jack thought that he was late then.

### 1.5 Sense without a referent

Early in "On sense and reference", Frege admits in unequivocal terms that sense without a referent is possible:

The expression "the least rapidly convergent series" has a sense but demonstrably has no referent, since for every given convergent series, another convergent, but less rapidly convergent, series can be found.  
(Frege 1892a: 159)

Gareth Evans has pointed out that many of Frege's subsequent discussions of the issue quickly introduce fiction, poetry, or in general non-serious uses of language. This led Evans to qualify the natural view that for Frege sense without a referent is straightforwardly an open possibility. As Evans put it:

Frege's later [i.e. post "On sense and reference"] apparent willingness to ascribe sense to certain empty singular terms was equivocal, hedged around with qualifications, and dubiously consistent with the fundamentals of his philosophy of language. (Evans 1982: 38)

Evans in part justifies the last remark, about the consistency with Frege's fundamentals, by the observation that if sense is glossed as the way in which an object is thought about, it would seem that there could be no sense without an object thought about, and so no sense without a referent. I read this as a reason for not regarding way of thinking, or mode of presentation, as a Fregean fundamental.

In "On sense and reference", Frege's willingness to recognize sense without a referent does not seem equivocal or qualified. He uses "Eigenname" very widely to include definite descriptions and whole sentences, in other words, to include semantically complex expressions as well as semantically simple ones, and the issue about sense without a referent takes a different turn for complex and simple expressions. Given Frege's attachment to compositionality principles, the sense of any complex expression ought to be built up out of the senses of its parts.<sup>2</sup> If this associates the complex with a way of thinking about an object, well and good; but the sense of the complex is assured, on Fregean principles, by the senses of the parts and their manner of combination, whether or not the whole constitutes a way of thinking of an object. If we take way of thinking or mode of presentation as Frege's central sense-investing feature for simple expressions, then indeed we are owed an explanation of how there can be a way or a mode in the absence of an object. But I have argued that we should drop modes of presentation as serious contributors to an understanding of Fregean sense. Instead sense features in various issues, like propositional attitudes (more specifically, in samesaying), in guaranteed sameness of referent (see essay IV: "Fregean sense"), in counting items of knowledge, and in saying which contradictions are manifest. But Evans's claim about what Frege actually said remains to be answered.

Frege discusses a simple proper name, "Odysseus", in "On sense and reference", saying that it is doubtful whether it has a referent (1892a: 162). The aim of his discussion at this point is to persuade us that if a sentence has a referent it is a truth value. He observes that if we take seriously the question whether or not a sentence is true, we take it for granted that each proper name it contains has a referent. This special connection between the referents of the parts and the truth value of the whole suggests that truth value is what should be counted as the referent of the whole.

Frege says that if "Odysseus" lacks a referent, then whatever the referent of

2 Frege takes various forms of compositionality for granted. But compositionality for natural language is quite a difficult issue, as argued in essay XI: "Two ways to smoke a cigarette".

a whole sentence containing it is will also be lacking, a remark which reveals his commitment to the compositionality of reference. So he asks what would be lacking in a sentence like “Odysseus was put ashore at Ithaca while still asleep” on the assumption that “Odysseus” lacks a referent, and answers that what would be lacking is a truth value. Hence the invitation to conclude that the sentence’s truth value, if any, is its referent, if any.

Lack of truth value only matters if we are engaged in serious questions of truth and falsehood, and does not matter if our only concern is with “aesthetic delight”. In the latter case, we can engage with thoughts, without considering whether they are true or false, for “the thought remains the same whether ‘Odysseus’ has a referent or not” (Frege 1892a: 163; 1906b: 191). Whatever one may think of the overall argument for truth values as the referents of whole sentences, there seems no equivocation in this discussion (nor in the 1906 discussion) about whether or not there can be semantically simple names which lack a referent. Evans’s view requires different evidence.

He expresses the positive view he claims to find in Frege in the following terms:

we may gloss those passages in which Frege says that a sentence containing an empty singular term may express a thought as follows. Yes: a sentence containing an empty singular term may have a sense, in that it does not necessarily have to be likened to a sentence containing a nonsense-word. But no: it does not *really* have a sense of the kind possessed by ordinary atomic sentences, because it does not function properly, it is only *as if* it functions properly. (Evans 1982: 30)

Evans here accuses Frege of confusing a situation in which it is pretended that an expression has a sense and a referent, though it does not really have either, with a situation in which an expression genuinely has a sense but lacks a referent. The passage Evans principally cites in support of this accusation is the posthumously published “Logic” dated to 1897, in which Frege says:

Names which fail to fulfil the usual role of a proper name, which is to name something, may be called mock proper names. (Frege 1897: 130)

Frege adds that a fictional sentence expresses a “mock thought”, a Scheingedanke. According to my soundings with native speakers, a Schein-*F* is normally something intended to seem an *F* even though it is not an *F*. On this reading, Frege is saying that a fictional sentence does not really express a thought, the explanation being that it contains an expression which looks like a proper name but which, because it has no referent, is not really a proper name.

However, I think that in this context Frege uses “mock” (or rather “schein”) in such a way that a mock-*F* is an *F* which is not to be taken seriously. For example, he says that thoughts in fiction “are not to be taken

seriously as in the sciences: they are only mock thoughts”; here “not to be taken seriously” seems precisely to be a gloss on “mock”. Since an *F* which is not to be taken seriously is an *F*, on this reading a Schein-*F* is an *F*. Reverting to Frege’s claim that “although . . . ‘William Tell’ is a mock proper name, we cannot deny it a sense”, we should understand it as affirming that the name has a sense just as much as does a name with a referent, and it is a mock name only in that it does not have a serious use.

A defender of Evans’s position may ask us to focus on this sentence:

Assertions in fiction are not to be taken seriously: they are only mock assertions. (Frege 1897: 130)

What appear to be assertions in fiction are not really assertions, for something that is not to be taken seriously (in the way Frege envisages) is not an assertion at all. Given that the very next sentence is one already quoted

Even the thoughts are not to be taken seriously as in the sciences: they are only mock thoughts

perhaps we should read the remarks in parallel: just as mock assertions are not assertions, mock thoughts are not thoughts (and the fact that in both cases “assertions” and “thoughts” occur in a way which might seem to delimit the subject matter can be dismissed as an unfortunate manner of speaking).

While it is true that a mock assertion is not an assertion, this is not because a mock-*F* is in general not an *F*. Evans’s opponent can properly object that a mock assertion is not an assertion because assertions are essentially to be taken seriously. A mock assertion would be one not to be taken seriously, so it would not be an assertion at all. This springs from a special feature of assertion, not from a feature of “mock”, so the result cannot be extended to mock thoughts. A non-assertive saying can express a thought, regardless of whether or not it is to be taken seriously. In this context, what it is not to take a saying seriously is not to take seriously the thought expressed, not to take it as something to whose truth the speaker commits himself. I conclude that this passage does not support Evans’s reading of Frege.

It may be that other passages are more on Evans’s side.<sup>3</sup> I will not attempt an overall assessment, but will discuss one passage which I find specially

3 There are some isolated remarks, for example: “The sentence ‘Leo Sachse is a man’ is the expression of a thought only if ‘Leo Sachse’ designates something” (Frege 1906c: 174). The Editors give reason for doubting their tentative attribution of the date 1906 to this fragment, and suggest that it may have been written in around 1880, well before “On sense and reference” (1892a). Certainly the text does not mention the sense–reference distinction explicitly. It would therefore be rash to count this as evidence for any vacillation on Frege’s part concerning sense without a referent.

interesting, and which introduces some fresh material: the comparison between empty names and “empty” predicates. By the time of his discussion of Schröder’s *Lectures on the Algebra of Logic* (Frege 1895), Frege has made it plain (as he did not in “On sense and reference”) that the referent of a concept-word (a predicate) is a concept. This makes room for the possibility that there are no intelligible predicates without referents, though there are predicates under which nothing falls. Frege reports that it is easy to confuse the two cases. However:

a proper name that designates nothing has no logical justification, since in logic we are concerned with truth in the strictest sense of the word; it may on the other hand still be used in fiction and fable. For concepts that do not comprehend anything under them it is quite different: they are entirely legitimate. (Frege 1895: 226)

If at this time Frege really had doubts about whether there could be sense without a referent, this would have been the place to air them. In fact, the point he is making is different: there is no place for empty names within logic, or more generally within any discipline concerned with truth (a science), for truth requires that names be non-empty.<sup>4</sup>

Even if the passage confirms Frege’s commitment to the possibility of sense without a referent, we can still explore the supposed difference between a name lacking a bearer and a predicate lacking satisfiers. Frege hints at more than one explanation. The first, though elliptical, seems to suggest that whereas there is normally a commitment on the part of one using a name to its having a referent, there is no parallel commitment on the part of one using a predicate to its having a satisfier. Put thus, the point about predicates seems incorrect. An assertive utterance of “John is happy” commits me to the truth of John being happy, and so as much commits me to “happy” having a satisfier as to “John” having a referent. We can use a predicate with no satisfiers to state serious truths, like that there is no round square (Frege 1895: 227), but the same goes for names: there is no such heavenly body as Vulcan. Frege moves to a second explanation: in using a name, its referent is normally something I think about, and wish to say something about; but it is not the case that in using a predicate its satisfiers are things I think about, or wish to say something about:

If I utter a sentence with the grammatical subject “all men” I do *not* wish

4 An exception to this rule which Frege notes, that an empty proper name may be used in a true ascription of a thought to another, again makes plain his firm commitment to the possibility of sense without a referent. No doubt he would not have regarded psychology (construed as including a systematic study of beliefs and other contentful mental states) as a science.



to say something about some Central African chief wholly unknown to me. (Frege 1895: 227)

This point, though still rather vague, seems to me correct, and to mark a contrast with proper names (as we use that term).

The possibility of sense without a referent seems to make the demand for some reductive account of sense more urgent. What can the sense of an empty name be? Not an object, since the name is empty; to say that it is given by a definite description takes us in a barren direction, one repudiated in §1.2 above. The view I favour is that the sense of a name is to be given through a specification, which will typically reuse the name in question, of the conditions under which it has a referent, for example: for all  $x$ , “Hesperus” refers to  $x$  iff  $x$  is Hesperus”. In contrast to McDowell, it is essential to the position I favour that such sense-revealing axioms do not commit the theorist to the existence of a referent. This means a departure from classical logic to a negative free logic, following a suggestion by Tyler Burge (1974). A biconditional axiom like “for all  $x$ , ‘Vulcan’ refers to  $x$  iff  $x$  is Vulcan” will have a right hand side (“ $x$  is Vulcan”) false of each object, and so will properly yield the result that “Vulcan” refers to nothing. The details are sketched in a little more detail in essay XII (“Sense without reference”).

It is not good enough just to put one’s foot down and insist upon homophonic specifications of sense within a free logic. There are at least two further tasks. One is to make a connection with the use of language, and this is attempted in essay XII. The other is to examine arguments directly favouring the view that sense without a referent is impossible, undertaken in essay IX (“Names, fictional names, and ‘really’”). There the target is Evans’s arguments in favour of what he calls the Russellian nature of typical proper names, where a Russellian name is one which owes its intelligibility to having a referent. If our ordinary names were Russellian, it would not be appropriate to describe them in terms of a semantic theory which made room for sense without a referent; but I suggest that Evans’s arguments in favour of the view that ordinary names are Russellian are unpersuasive. The points apply with equal force to the essentially Russellian position which Evans ascribes to Frege.

### **1.6 The determination of reference, sharp boundaries, saturation**

Sense determines referent at least in that expressions with the same sense have the same referent. Some have thought that Frege was interested in a stronger relation of determination. For example, Dummett moves swiftly from the point that, for Frege, sense must be something more than reference, to the claim that, at least in an initial simple model, the sense of a name corresponds to a distinctive way of recognizing an object as the referent of the

name (Dummett 1973: 98). Applying this to whole sentences, whose senses are thoughts, it would seem to follow that a true thought corresponds to a distinctive way of recognizing the true as its referent, and a false one to a distinctive way of recognizing the false as its referent. In short, it would appear that understanding a sentence, that is, grasping its sense or knowing which thought it expresses, would involve being able to recognize it as true, if it is true, or as false, if it is false. Dummett's initial presentation of Frege casts him as some kind of verificationist, though, as Dummett himself later stresses, concerning one formulation of verificationism, there is no evidence that Frege was tempted by such a position.<sup>5</sup>

Dummett discusses various ways of saying in what knowledge of sense would consist. One way involves the notion of *recognition* of the referent, another that of *knowing what it is for something to be* the referent of an expression. He has some sympathy with the former, since recognitional knowledge need not be verbalizable, whereas he affirms that "we can give little content to the notion of possessing knowledge of this [latter] kind unless it be verbalizable knowledge" (Dummett 1973: 231). Later he suggests that to grasp an expression's sense "is to have an implicit understanding of a general rule which constitutes the contribution made by that expression to what has to be done to determine any sentence in which it occurs as true or as false" (Dummett 1973: 236), and there are versions which turn on the notion of knowledge of a criterion. In one form or another, views similar to that which Dummett ascribes to Frege are common. Thus Gareth Evans, though officially drawing his inspiration at this point from Russell, thought that one who understood a name should have the capacity to distinguish its bearer from all other things. Positions of this kind appear to be at variance with the very weak conditions that we in practice place upon what understanding a name involves. Someone in a conversation starts using a name which is unfamiliar to me, but it takes little time for me to count as a party to this name-using practice, even though I have no independent capacity to distinguish the bearer from all other things: the best I could do would be to describe the bearer as whoever the other person was referring to by the name.

This undemanding condition on understanding names harmonizes with the pared down Fregeanism recommended here: there is no describing in terms of an independent identificatory capacity what knowledge of sense consists in. There is no need to aim for such a description, for there is no ignorance which it would ameliorate. To understand "Hesperus", to know or grasp its sense, is to know that it refers to something iff that thing is Hesperus. More

5 The formulation is "for every true statement, there is something which we could know and which, if we knew it, would establish conclusively for us the truth of the statement". Dummett comments: "What Frege's stance would be on this thesis, there is nothing in his writings to determine" (Dummett 1973: 231).

elucidation of what this knowledge “consists in” receives only an externalist answer, rather than an answer in terms of the thoughts the knower has: one possesses the knowledge by being party to the practice of using “Hesperus” as a name which is supposed to refer to Hesperus. Attention shifts, accordingly, to the notion of a name-using practice, and this is constrained by causal links among practitioners (see essay XII: “Sense without reference”). Although coordinating intentions doubtless play an essential part in binding people into a practice, a further essential part is played by causal links of which the participants need have no awareness. These links could accordingly not be formulated within the kind of framework Dummett proposes, as things implicitly or explicitly known. In general, facts which are not known may be essential to the possibility of knowledge.

There is another way to see the determination of referent by sense as stronger than the merely minimal relation endorsed so far. One may see an expression’s referent as a function of two elements: its sense, and the way the world is. If “Fido barks” is true, this is in part due to the sense of the sentence, and in part due to how things are. Had things been different in either respect, the sentence might not have been true. In this way, the sense/reference distinction, applied to sentences, can be seen as introducing a distinction between two distinguishable contributors to truth or falsehood, sense itself being one, and the other being whatever must be added to sense to deliver the referent, a truth-maker or a falsehood-maker. The analogy can be extended to subsentential expressions: the sense of the name “Hesperus”, together with various astronomical facts, ensures that its referent is what it in fact is. Various linguistic phenomena, for example vagueness, show that this is far from a truistic picture. On the face of it, a vague expression, be it a proper name, a predicate or a sentence, can have a sense yet, even given how the world is, fail to determine as referent a determinate object, concept or truth value. And some have argued that features of language other than vagueness generate something similar (see the discussion of the “Travis effect” in essay XI: “Two ways to smoke a cigarette”). More fundamentally, the picture of dual determination presupposes that sense could remain constant through transplantation to a different environment, and this is inconsistent with even mild forms of semantic externalism.

One reason for which Frege held that concepts must have sharp boundaries is that he thought of them as functions whose values are truth values. Assuming, as Frege did, that there are just two truth values, there is simply no room for vagueness. Indeed, even infinitely many values arguably cannot model vagueness (as I argue in essay III: “Concepts without boundaries”). Moving from complete to partial functions will not address the fact, or apparent fact, that there is no sharp boundary to the borderline cases of vagueness. Frege himself did not appeal to this point in explaining his insistence upon sharp boundaries; rather, he appealed to their importance to inference. His remarks are sketchy and hard to interpret (cf. Frege 1896; for a discussion see

Diamond 1984). However the view that classification is a matter of drawing boundaries is widespread. In essay III I sketch an alternative picture.

The completing of the essentially incomplete function with its argument, an object, connects with a puzzle that troubled Russell. In Frege's terminology, the puzzle is: what would it be for an object to *saturate* a function, in particular a concept, as opposed to merely coexisting with it? In Russell's terminology, the problem is that of the unity of the proposition, which one might express in this way: what is it for an object and a universal to combine in such a way as to produce a *proposition*, as opposed to merely coexisting? Frege does not express puzzlement; correspondingly, nothing he says, in my opinion, is offered as a solution to a puzzle. For example, in "On concept and object" he is explicit that

What I here call the predicative nature of the concept is just a special case of the need of supplementation, the "unsaturatedness", that I gave as the essential feature of a function [in "Function and concept"].

(Frege 1892b: 187).

Clearly he did not think that any puzzle there might be about how something could have a predicative nature could be solved by calling it unsaturated.<sup>6</sup> Russell, by contrast, struggled with the problem, as I describe in essay V ("How can some thing say something?"). I think that Russell was right to think there is something problematic here, but that at least one version of the problem is resolvable. Let us use "concatenation" for whatever the allegedly problematic union is between function and argument, or object and universal, thanks to which we end up with a single complex entity, a thought or a proposition, fit to have a truth value. If we can say in full detail what concatenation is and how it works, we have said all that it is reasonable to ask about how there could be such a relation. And we can say what concatenation is by adapting the kind of recursive truth theory which Davidson developed from Tarski. To concatenate the object 7 and the concept of being prime is to bring about something which is true iff 7 is prime; and so on for all the sentences of the language.

### **1.7 The nature of sense**

To construe the doctrine that sense determines reference as no more than the claim that expressions with the same sense have the same referent is consonant

6 This opinion may differ from Dummett's, who writes that, for Frege: "the problems [of the unity of the proposition] are entirely spurious", thanks to the Fregean insight that "it is of the essence of a concept or property to be *of* an object, of a relation to be *between* two objects" (Dummett 1973: 175, 176).

with the modest homophonic way of specifying senses recommended in §1.2 above. However, all this modesty may be somewhat frustrating, and constantly saying what sense is *not* may create the feeling that the nature of sense cannot be positively characterized at all. In fact, as already mentioned in passing, sense receives a positive characterization through its roles: counting items of knowledge, characterizing samesaying and manifest contradiction, being involved in the correctness of reports of speech, and providing a guarantee of sameness of reference. The last role is discussed in essay VII (“Fregean sense”). The idea is that we can take it as a datum that distinct tokens can sometimes unthinkingly, but with full justification, be taken as having the same referent. This is what Frege envisaged would happen with identity sentences “ $a = a$ ”. This is not a matter of their form. The hypothesis that Aristotle the philosopher was reincarnated as Aristotle the shipping magnate could be expressed as “Aristotle is Aristotle”, in which case it would be false, and one who took the occurrence to be trivially true would be mistaken. It is not that, for the trivial cases, we apply antecedent knowledge of sameness of sense to derive the trivial truth; rather, the fact that on occasion we are fully justified in taking sameness of reference for granted is what marks sameness of sense.

To round out this account, something needs to be said about the circumstances under which this justification arises. I suggest that one case in which it does so is when there is a single name-using practice in which a speaker and a hearer are participating. Then it is right for the hearer to take for granted that this occurrence of a name has the same referent as the referent of his own use of an equiform token in the same context. This is what makes it possible for hearers to reuse a name in their reports of what has been said. A name-using practice, in turn, is in part determined by causal links in rather the way that Kripke suggested, though without Kripke’s commitment to a referent (which is argued to be gratuitous). The suggestion is spelled out in essay XII (“Sense without reference”).

The need to bring the two ideas together (sameness of sense as a guarantee of sameness of referent, and senses of names individuated by name-using practices) becomes plain when we consider the unhesitating way in which we take for granted that even perfect strangers who speak our language are in fact doing so. We can understand, and so come to know what is said, without any independent evidence that the speaker is using words as we would, or even that he is intending to do so. This is a datum for any descriptive account of the epistemology of meaning, that is, of any account of understanding. One hypothesis about the underlying metaphysics is that in one way or another our use and the stranger’s use have common sources just as, in the Adam and Eve story, the stranger and I have common ancestors. What makes it correct to take community of understanding, and so of referent, for granted is some causal coordination, which it would be very hard to spell out in detail, but which there is no need to suppose involves any mental facts

about the participants, beyond their possession of the relevant knowledge in these cases, other than their intentions to refer to what the others refer to.

This already takes an externalist step, and one possibly not congenial to the historical Frege. I see no tension between such an epistemology and the enduring aspects of Frege's thought. In essay VI ("Easy possibilities") I raise the question of what sort of epistemology would allow us to say that we do indeed understand our words despite the fact that we might well not be sensitive to small changes in their meaning. This question is addressed to Williamson's version of the epistemic theory of vagueness, but the externalist account of the species of knowledge which constitutes understanding is to be recommended more widely.

### **1.8 An overlooked position**

A caricature of quite a body of contemporary opinion is that we have to choose between an essentially Millian position about proper names (and, perhaps, other singular terms) according to which there is nothing more to their semantics than their referent, and a so-called Fregean position, according to which all singular terms are essentially descriptive. Debate oscillates between the positions, with arguments against one being offered as arguments for the other, and much defensive work devoted to refinement, dealing with apparent counterexamples, and so on.

Pared down Fregeanism offers an overlooked option. It shares with Millian views a denial of descriptivism, and with conventional Fregean views the admission of the possibility of coreferring expressions which require different semantic descriptions, and the possibility of names which lack a referent. Accordingly, it holds out the prospect of having all the virtues of the current major approaches without any of their difficulties.

## **2 The essays: provenance and comments**

The essays are arranged in approximately the chronological order of their publication. They are reprinted with only very minor changes in the text (except in the case of essay XI). Where it was clear to me that the originals were inadequate in a way that could be addressed in a brief note, I have added one, marked in a distinctive style from the original notes.

I "Understanding and theories of meaning". *Proceedings of the Aristotelian Society* 80, 1979: 127–44. Presented at a meeting of the Aristotelian Society in Tavistock Place, London WC1, on Monday 3 March, 1980, at 6:30 p.m. Reprinted by courtesy of the Editor of the Aristotelian Society: © 1979.

The main points are that revealing compositionality is a virtue in a theory of meaning that cannot be derived from other virtues (like that of stating

something sufficient for understanding); and that Davidson's account of radical interpretation can say what it is for a truth theory to be interpretive iff it could say what it is for a meaning theory to be correct, where a meaning theory is one whose focal concept is *means that* rather than *is true iff*.

The background is Davidson's view, in the 1970s, that a theory of meaning for a language should state something knowledge of which would be sufficient to enable the knower to achieve understanding of the language with which the theory deals. In this context, a theory is a set of sentences closed under the relation of logical consequence. The sentences belonging to the set are the theorems of the theory, and theorems which do not belong to the set in virtue of the closure relation are axioms. Knowing a theory is not the same as knowing that all the theorems are true: I might know this on authority, while having no idea what the theory says, and one would want monolingual speakers of different languages to be able to know the same theory. Knowing a theory is, approximately, knowing, concerning what the axioms state, that things are thus (i.e. as the axioms state them to be), and being able to use this knowledge to come to know, subject to limitations of computational and memory capacities, concerning what any theorem says, that things are thus (i.e. as the theorem states they are).

Davidson's favoured form for a theory of meaning is a truth theory, or "T-theory" for short, one which contains, for each sentence of the object language, a "T-sentence", that is, a theorem of the form "*s* is true iff *p*". He was clear that if there are any true truth theories for a language, there are some knowledge of which would not be sufficient for interpretation. For example, if there is a true T-theory whose axiom for "green" is modelled on the claim that something satisfies "green" iff that thing is green, there is a true T-theory whose axiom for "green" is modelled on the claim that something satisfies "green" iff that thing is green and is either square or not square; but this second T-theory will not be good for arriving at correct interpretations of sentences containing "green". Moreover, the closure relation  $\vdash$  in classical theories has the following feature:

$$\text{If } \vdash P \leftrightarrow Q \text{ and } \vdash R \text{ then } \vdash P \leftrightarrow (Q \ \& \ R)$$

which ensures, on an obvious further assumption, that if "*s* is true iff *p*" is a theorem, so is "*s* is true iff (*p* & (*q*  $\vee$   $\neg$  *q*))", and the latter is not propitious for the work of interpretation.

In the 1970s, Davidson distributed the knowledge which would suffice for interpretation into two parts: there was the theoretical knowledge, knowledge of a theory (in the current restricted sense of "theory"), but to this was added an extra-theoretical or reflexive component, containing information about the theory itself, or about what the theory stated. This second part would select some theoretical knowledge as "canonical", roughly speaking, as derived with minimum deductive fuss, in this way addressing the second

problem mentioned in the previous paragraph; and it would select some theories as “interpretive”, that is, as properly grounded in the canons of radical interpretation, thus addressing the first problem mentioned in the previous paragraph. The final claim, at this stage in Davidson’s position, was that the following knowledge would be sufficient for interpretation: to know a T-theory, to know that it is interpretive, and to know how to filter out any non-canonical theoretical knowledge.

This split view is one main topic of the essay reprinted here; the other is the proper justifications for various virtues that a theory of meaning may possess. On the last topic, it is argued that compositionality is an independent virtue, and not one which can be derived from other features which Davidson prized. Reverting to the first topic, it is argued that the relevant reflexive knowledge must take a form which causes it to undercut the need for the theoretical knowledge. All that is required is just pairings of sentences with states of affairs, together with knowledge that these pairings were established in the appropriate way on the basis of evidence. The pairings need not be organized into sentences at all, let alone into true sentences, let alone into sentences known to be true. Any such further condition would be gratuitous, since just access to the pairings (or more exactly to the mechanisms which generate them, analogous to T-theoretic axioms), together with knowledge from which one can infer that they can be used for interpretation, is enough. The upshot is that however important the notion of truth is in the reflexive component (where it features in principles like charity, which constrain interpretation), a central role in the other component is not justified by Davidson’s methodology.

Summing up the reflections (derived from John Foster 1975) which led to the split view, Davidson (1976) concludes: “nothing strictly constitutes a theory of meaning”. This is because a theory of meaning is, by definition, both a theory and also epistemically sufficient. Once an extra-theoretical or reflective component is necessary, no theoretical knowledge is sufficient. By 1984, when the Introduction to his *Enquiries into Truth and Interpretation* was published, he was happy to summarize the extra-theoretical knowledge required to establish a sufficient condition as simply that the theory is nomic (1984: xviii). This suggests that he could instead have adopted a non-split view, in which purely theoretic knowledge is epistemically sufficient. One could enrich the metalanguage with an operator with the meaning “it is nomic that:” and require that the theory prove, for every object language sentence, a theorem of the form “It is nomic that:  $s$  is true iff  $p$ ”. If the extra-theoretic knowledge that the theory is nomic yields an epistemically sufficient condition, why should not intra-theoretic knowledge of nomically qualified theorems not achieve the same?

How the suggestion would deal with the difficulties raised here cannot be assessed without some assumptions about the logic of “it is nomic that”. One natural view is that it is a restricted version of logical necessity, so that “it is necessary that P” proof-theoretically entails “it is nomic that P”, and if we



have a proof of nomic equivalence, we also have a proof of anything which results from substituting one of the equivalents for the other within any theorem, provided that the substitution does not occur within an operator stronger than that of nomic necessity. If so, our difficulties remain, for the necessary equivalence between “ $x$  is green” and “ $x$  is green and either square or not square” will be enough for their nomic equivalence, so having nomic axioms will not ensure having interpretive ones. Likewise, even if a theory has interpretive T-theorems, it will also have uninterpretive ones, by the addition of unwanted conjuncts to the right hand side of the biconditional. Addressing the second difficulty in terms of canonical proof still leaves a split view of what is epistemically sufficient: we would need to know the theory, that is, the facts the theory states, and also need to know which of these are expressed by sentences which have been derived in one rather than another way.

II “Evans on reference”. Originally published as R. M. Sainsbury, “*Critical Notice: The Varieties of Reference* by Gareth Evans”, *Mind* **94**, 1985: 120–42. Reprinted by kind permission of the Mind Association.

This essay was intended to serve as an introduction to Evans’s book *The Varieties of Reference*. It is reprinted here with the same hope.

The account is mostly descriptive, the main criticism being of Evans’s preference for a very demanding account of mental reference, according to which to think about an object requires the capacity to distinguish it from all other things. He draws this from Russell, but he could as well have drawn it, if not from Frege himself, at least from Dummett’s Frege.

Evans criticizes an alternative and less demanding view, which he calls the Photograph Model, but I raise doubts about his argument.

III “Concepts without boundaries”. London: King’s College London, 1990. This is the text of the philosophical part of my inaugural lecture as Susan Stebbing Professor of Philosophy at King’s College London, delivered on 6 November 1990; hence the absence of references and footnotes. It was reprinted in Rosanna Keefe and Peter Smith (eds) *Vagueness: A Reader*, Cambridge MA: MIT Press, 1996: 251–64.

This essay challenges Frege’s claim that concepts draw sharp boundaries, which appears not to fit well with the majority of our concepts, for these are vague. I suggest that these do not draw boundaries at all, and so it is literally false that there is such a set as, for example, the set of red things.

Rather than engage with the basis for the Fregean position, I try to describe an alternative, one in which concepts are thought of not in terms of boundaries or edges which divide up logical space, but rather in terms of poles which mark central cases around which other cases cluster. These are concepts without boundaries.

The difference between the positions can be brought out in terms of the role of negation. Within the classical position, grasp of a concept automatically constitutes grasp of its negation: it involves a conception of the boundary separating what the concept includes from what it excludes. So grasping “red” involves knowing not only what is involved in being red, but also what is involved in not being red. On the alternative conception of concepts as without boundaries, grasp involves an appreciation of central cases, and also an appreciation of central cases of competing concepts. Grasping “red” involves some appropriate contact with a central case (being able to recognize something as red, or imagine it to be red) together with a positive appreciation of competing concepts in the area, for example “orange”. This explains why the most conspicuous vague predicates come in groups of contraries, and it also makes room more readily than competing accounts for the fact that it is quite legitimate to use such concepts in drawing temporary boundaries (“for present purposes, ‘excellent’ means 75 per cent or above”), without supposing either that one is behaving inconsistently with the concept or that one is following its predetermined contours.

IV “Russell on names and communication”. In Andrew Irvine and Gary Wedeking (eds) *Russell and Analytic Philosophy*, Toronto: University of Toronto Press, 1993: 3–21. Reprinted with permission of University of Toronto Press. The volume collects papers presented at a conference on Russell held at the University of British Columbia, Vancouver, in 1992.

Kripke (1972: 27) said that both Frege and Russell held that “a proper name, properly used, simply was a definite description abbreviated or disguised”. It is unclear whether Frege was committed to this generalization. Although he famously used definite descriptions in providing an example of how opinions as to the sense of “Aristotle” may differ (Frege 1892a: 158n4), it is not obvious that one is supposed to generalize. His account of indexicals certainly entails that some kind of non-descriptive reference is possible, for he holds that “the time of utterance is part of the expression of a thought” (Frege 1918: 358), and whatever a time is, it is not a description.

Russell’s views are much easier to discover, and are, at least in *Problems of Philosophy* (1912b), blatantly not as Kripke describes.<sup>7</sup> Kripke takes it that Russell thought that descriptions could be used to specify some kind of public

7 In “A puzzle about belief” Kripke qualifies his attribution of a description theory of names to Frege (Kripke 1979: 271n3), and suggests that it forces Russell to hold that users of a name speak different languages (“idiolects”: Kripke 1979: 277n29). That is partly right (the associated descriptions vary from speaker to speaker), but does not do justice to Russell’s pointers towards a public and shared language. It would, of course, be absurd to suggest that, even if I am right about these exegetical issues, they in any way detract from Kripke’s philosophical contribution.

meaning or use for proper names (as this phrase is ordinarily understood). In fact, in the period in question, Russell was not much interested in questions about public language, and to the very modest extent to which he did try to account for the role of proper names in public language, it was in a way which made only their referent relevant. It is ironic that although Russell's theory of descriptions and of names features in almost every philosophy of language course, Russell himself did not see these theories as contributions to the philosophy of language, as opposed to the philosophy of judgement. In Russell's perspective, the difference is huge: judgement is based on acquaintance, and no two people can be acquainted with the same individuals, so no two individuals can share a singular judgement. By contrast, a public language is a shared language. This is a gap which could not open up within Frege's position, for thoughts or judgements are senses, and, setting aside possible complications related to indexicality, anyone can in principle access any sense.

The view which, I argue, Russell really advances is that definite descriptions can be used to give an account of what is going on in the mind of one who uses a name, and the right description may differ from speaker to speaker, and even for a given speaker from occasion to occasion. That is where the account of judgement stops. There is a brief sketch of the way in which communication can take place, despite these differing judgements. What counts for truth conditions and for success in communication are not a name's associated descriptions but their denotation, that is, the name's bearer. The result is a two-level view, of a kind which has been popular recently (versions of it are found in Evans 1982 and Récanati 1993). The unit of judgement is descriptive, but the unit of communication is what would now be called a Russellian proposition.

V "How can some thing say something?". In Ray Monk and Anthony Palmer (eds) *Bertrand Russell and the Origins of Analytic Philosophy*, Bristol: Thoemmes, 1996: 137–53. Reprinted by permission of Thoemmes Press. The volume collects papers from a conference on Russell held at the University of Southampton in July 1995, which included a symposium between Stuart Candlish and myself. Without informing me, the editors published my paper under a title of their choosing ("How can we mean something?").

Under the heading of "the unity of the proposition", Russell raised a number of distinguishable questions relating to the way in which words, or their non-linguistic correlates, could combine to form the kind of unity which is capable of saying something. My essay supports the judgement that there is a genuine puzzle here, and one which arises within frameworks other than Russell's. Within Frege's system, one question is what it is for an argument to saturate a function so as to deliver a value, rather than for argument and function simply to coexist.

I suggest that to the extent that we have an intelligible question, we also have an answer, briefly described at the end of §1.6 above, and argued for in

more detail in the essay: the answer is given by a recursive specification of truth conditions, in the style which Davidson has developed from Tarski.

VI “Easy possibilities”. *Philosophy and Phenomenological Research* 57, 1997: 907–19. Copyright © International Phenomenology Society 1997. Reprinted with permission.

The main idea is to apply the reliability conditional (“If you know, you couldn’t easily have been wrong”) to knowledge of meaning, bearing in mind the plausible view that shifts in the meanings of our words could occur undetectably. Knowledge in some ways resembles reliability (a reliable machine is one which could not easily go wrong) and safety (a safe action is one which could not easily have resulted in disaster). These notions require margins for error: for example, an action may be unsafe even though no disaster ensued because it did not leave enough margin. And margins for error are antithetical to KK principles (if you know, you know that you know) or their analogues for reliability and safety.

In his book *Vagueness* (1994a), Timothy Williamson argued that we (by and large) know the meanings of our words, yet they could have meant something slightly different while our verbal behaviour remained largely unchanged. There is an apparent tension, for it sounds as if the meanings of our words could easily have been different, in which case it appears to follow that we could easily have been wrong about their meaning, which, via the reliability conditional, entails that we do not in fact know their meaning. The essay attempts to show that there is no inconsistency. The crucial issue is what makes a possibility “easy”. In answering this question, one needs to bear in mind that one could easily have failed to know things one in fact knows, and that this is consistent with the reliability conditional.

The essay was originally intended to be one of several short discussions of Williamson’s book, *Vagueness*, but under Ernest Sosa’s helpful influence it became a longer piece which was published separately.

VII “Fregean sense”. In Timothy Childers, Petr Kolár and Vladimír Svoboda (eds) *Logica ’96*, Prague: Filosofia, 1997: 261–76. The volume collects papers presented at the annual Logica conference held at Liblice Castle in the Czech Republic, June 1996.

This essay develops an idea derived from Peter Strawson and also, in a more detailed form, John Campbell, that Fregean sense should be thought of as guaranteed sameness of referent. It also discusses some criticisms which Ruth Millikan has addressed to Frege and Fregeans. A caricature of her target would be that explaining our fallible capacities for recognizing sameness and difference among ordinary objects somehow presupposes infallible capacities for recognizing sameness and difference among senses. I accept Millikan’s

main positive points, and they should make Fregeans aware of the danger of hoping for more explanatory value from sense that it can deliver. However, a suitably pared down Fregeanism has no commitment to the positions which Millikan attacks.

A central point of the essay is that we cannot even make sense of the notion of formal validity (validity in virtue of logical form) without accepting something like Frege's distinction between sense and reference.

VIII "Indexicals and reported speech". *Philosophical Logic: Proceedings of the British Academy* 95, 1998: 45–69. Edited by Timothy Smiley. Reprinted by permission of Oxford University Press. The volume collects papers presented at a British Academy conference on philosophical logic on 16 March 1997, organized by Professor Timothy Smiley. Jimmy Altham was the co-symposiast.

This essay explores the hypothesis that we can throw light on Fregean sense as what is common to a correct report of speech and what is reported. The hypothesis is uninteresting when indexicality is bracketed: it merely has such consequences as that two occurrences of "Nothing travels faster than light" have the same sense. It is more interesting when indexicals feature in the speech reported. I must report Jack's "I am hungry" as Jack having said that he is hungry. The essay explores the consequences of this idea for truth theories, for the notion of "indexical thoughts", for the proper expression of scientific theories, and for the view that there are things which even God could not know.

It is argued that indexical expressions are dispensable in the content part of a report of speech, and this has implications for the extent to which indexicals are essential in explanations of action. The fact that one person may not be able to utter words which make her and another samesayers without doing more than this does not entail that there is no way in which she can samesay with the other. If yesterday Jack utters "That's a rabbit" in the woods, there may be no words that Jill can use today to say just the same without also saying more. But, saying more, she can samesay with Jack by saying he pointed at something yesterday in the woods and said that it was a rabbit. The approach reveals why a Fregean should put no faith in "modes of presentation": Jill's "it" cannot in any natural way be said to invoke the same mode of presentation as Jack's "That", yet on the assumption that the word is involved in a correct report of what Jack said, other Fregean principles suggest that Jill's "it" and Jack's "that" have the same sense.

IX "Names, fictional names and 'really'". *Proceedings of the Aristotelian Society, Supplementary Volume* 73, 1999: 243–69. Reprinted by courtesy of the Editor of the Aristotelian Society: © 1999. This comes from a symposium at the July 1999 Joint Session of the Mind Association and the Aristotelian Society. David Wiggins was my co-symposiast: hence the invitation at the end of my paper for him to expand his views.

Although it has been widely held that proper names and other singular terms require a referent for their intelligibility, arguments for this opinion are less easy to discover. Considerations which supposedly favour the need for a referent are often nothing more than reasons for dissatisfaction with descriptivist theories. These would constitute reasons for referent-involving theories only if there were no other options, but the view I favour amounts to such an option: there can be sense without a referent, and it can be specified in the modest homophonic way (within a free-logical framework).

Evans is relatively unusual in providing direct arguments for the need for a referent, though this essay claims that they are unconvincing. He also provides a ground-breaking attempt to make room for empty names by appealing to pretence. One aspect of this is the “game-to-reality shift”, which is involved in the use of “really”, whereby Evans seeks to account for genuinely true singular denials of existence. I argue that his account of “really” is incoherent, and that an alternative account, offered by David Wiggins, remains at best unfinished. If sense without a referent is allowed, approaches to existential claims which depend upon “really” become otiose.

X “Knowing meanings and knowing entities”. First published in Uxe Meixner and Peter Simons (eds) *Metaphysics in the Post-Metaphysical Age: Proceedings of the 22nd International Wittgenstein Symposium, 1999*, Vienna, 2001: 106–15. Copyright © öbvahpt Verlagsgesellschaft, Vienna, 2001. Reprinted with permission.

There is no doubt that, for Frege, senses are entities, and that understanding an expression is a matter of standing in some relation (“grasping”) to the entity which is its sense. Under Quine’s influence, such views became unpopular in the 1960s and 1970s; but they are currently re-emerging. The essay argues that simple-minded attempts to describe understanding in terms of a relation to an entity lead to manifest falsehood, and that in more complicated attempts the entities drop out of view: what becomes important is not the identity of the entity, but how it is thought of.

For traditional Fregeans, this poses a problem. But it is unproblematic for the pared down Fregeanism I advocate, for that view abandons senses as entities, and characterizes knowledge of meaning in terms of participation in a practice.

XI “Two ways to smoke a cigarette”. *Ratio* 14, 2001: 386–406. Reprinted with the permission of Blackwell Publishing. The original version of this paper was published alongside two papers on compositionality, one by Paul Horwich (2001) and one jointly by Jerry Fodor and Ernest Lepore (2001), and contained a discussion of their disagreement. The present reprinting excises that material and makes some consequential adjustments.

Frege was firmly committed to compositionality for both sense and reference. In considering natural language, however, while it is truistic that we understand sentences thanks to understanding the words which make them up, it is not obvious whether understanding the words gives us all the required information, or whether we must also bring to bear non-semantic information to arrive at a reasonable interpretation of the sentence. For example, the first noun relates differently to the second in the pairs “vacuum cleaner” and “carpet sweeper”. Does an appreciation of this difference belong to understanding? If so, it may be difficult to see how it can be extracted from knowledge of the meanings of the parts, and compositionality is threatened.

This essay suggests that there is no apriori route to resolving such issues, which require detailed attention to a range of specific cases. The kind of solution most often favoured in the essay is to see meaning or sense as relatively unspecific, and what are sometimes called different “readings” as merely different ways in which a single determinate sense can be made true. For many of the cases discussed, the preferred analogy is with the fact that “John ran” can be made true either by his running east or his running west, rather than with the fact that “John went to the bank” can be made true either by his going to a river bank or to a financial bank.

XII “Sense without reference”. In A. Newen, U. Nortmann and R. Stuhlmann-Laeisz (eds), *Building on Frege*, 2001: 211–30. Stanford, CA: CSLI Publications, 2001. Copyright © 2001 by CSLI Publications. Used with permission of CSLI Publications. The volume collects papers delivered at a conference on Frege in Bonn in October 1998.

This essay sketches a pared down Fregeanism which allows for sense without a referent. The bulk of the paper suggests an account of our use of names which does not discriminate between empty and non-empty ones. If this is right, then a semantics which likewise does not discriminate is appropriate. The connection with use will, I hope, address the worries of those who fear that although we can appeal to the referent of a non-empty name in individuating the practice of its use, we clearly cannot do the same for empty names. Instead, I suggest that name-using practices are individuated by their sources. The pattern is essentially like Kripke’s save that, I suggest, the referent, even if there is one, need play no individuating role.

For suggesting that I should reprint these essays, for guiding me in their selection, for judicious criticism, and for a great deal of support and encouragement, I am much indebted to José Luis Bermúdez: I greatly appreciate his help. I would also like to thank the Leverhulme Trust for a Senior Research Fellowship, during the tenure of which I prepared the essays for publication and wrote this introduction.

# Understanding and theories of meaning

---

The expression “theory of meaning” can be used in at least these two distinct ways:

A theory<sub>1</sub> of meaning relates to a single language, L, and purports to state, or in some other way to fix, the meaning of each L-sentence.

A theory<sub>2</sub> of meaning relates to language in general, and attempts to analyze, elucidate, or determine the empirical content of, the concept of meaning in general.

The issues I wish to discuss lie within the hypothesis that a theory<sub>2</sub> of meaning should essentially involve consideration of theories<sub>1</sub> of meaning. But this terminology is cumbersome. I shall reserve “theory of meaning” for theories<sub>1</sub> of meaning. Reformulated, the hypothesis is that a philosophical account of meaning should proceed by considering the nature of theories of meaning.

The hypothesis has initial plausibility, for it might seem that if theories of meaning specify meanings for particular languages, to reflect on the general nature of theories of meaning is generally to reflect on what each specifies, and is thus to reflect on the general character of meaning.

I do not seek to destroy this framework, but only to challenge certain contentions that have been made within it. I shall assume that, in the present discussion, our only ultimate philosophical goal is to attain a philosophical understanding of the nature of meaning in general. What I shall challenge is either the justification for or the correctness of certain constraints on theories of meaning that, it has been proposed, are required in the light of this ultimate aim. The constraints are these:

- (i) A theory of meaning for L should exhibit the semantic structure of L;
- (ii) A theory of meaning for L should have a syntax no less extensional than the syntax of L;
- (iii) A theory of meaning’s pairing theorems should not use as filling the expression “means that”, or cognates like “can be used to say that”.



[The “pairing theorems” are those which are supposed to state, or in some other way to fix, the meanings of the object language sentences, and I shall assume that they have the syntactic form “ $s \dots p$ ” where what replaces “ $s$ ” designates a sentence, whose content is supposedly given by the sentence in use which replaces “ $p$ ”; the “filling” is what fills the dots.]

The following three sections take up these claims in turn. All three constraints are satisfied by identifying a theory of meaning with a truth theory of the kind that Donald Davidson (1967, 1973) has proposed. Thus, much of what I argue for could be expressed: it is a mistake to think that truth theories, or philosophical reflection thereupon, can assist a philosophy of meaning in the way often supposed in the Davidsonian tradition.

## I Understanding and structure

If a theory of meaning for  $L$  exhibits the semantic structure of  $L$  it does so by identifying  $L$ 's semantic elements and deriving the pairing theorems relating to  $L$ -sentences from axioms relating to (typically subsentential)  $L$ -elements. In this section I consider some, but not all, of the ways one might attempt to justify the requirement that theories of meaning should exhibit semantic structure. In particular, I shall consider how one might attempt to derive it from two claims about the nature of theories of meaning, claims which would link the theories with the philosophy of meaning:

- (a) A theory of meaning for  $L$  is a theoretical representation of a practical ability possessed by anyone who understands  $L$ : that of recognizing the force and content of suitably presented  $L$ -utterances.<sup>1</sup>
- (b) A theory of meaning for  $L$  states something knowledge of which is sufficient for understanding  $L$ .<sup>2</sup>

(Those who define “theory of meaning” so as to exclude theories which do not exhibit semantic structure can regard the issue of this section as how to justify considering theories of meaning (in their sense) rather than other kinds of meaning-specifying theories.)

1 Michael Dummett (1976: 69) characterizes a theory of meaning as a theoretical representation of the practical ability of knowing how to speak the language. John McDowell identifies understanding a language with the recognitional ability I invoke: “Understanding a language consists in the ability to know, when speakers produce utterances in it, what propositional acts, with what contents, they are performing” (1976: 45). McDowell imposes a structural constraint, but does not attempt to derive it from the requirement that a theory represent the recognitional ability.

2 This claim is implicit in Davidson’s “Truth and Meaning” (1967: 29) and explicit in his “Radical Interpretation” (1973: 125). Cf. also the editors’ introduction to Evans and McDowell (1976) and McDowell (1977: 159).

Bending our thoughts to (a), let us consider the nature of recognitional abilities in general. One who can under certain circumstances recognize objects of kind K as objects of kind K is one who can, under those circumstances, come to know of any such object, suitably related to the circumstances, that it is of kind K. Recognition relates to kinds or types, and is explicitly or implicitly relative to appropriate circumstances, typically those of sensory presentation. Thus one who can, just by looking, recognize *Amanita Phalloides* as *Amanita Phalloides* is one who can, when any mushroom of this kind is close enough within his visual range, come to know that it is an *Amanita Phalloides*. This exemplifies a shallow representation of the ability, since the very concepts which occur essentially in the knowledge which the recognitional ability is the ability to acquire also occur, accidentally, in the description proffered of the kind of objects recognized. A deeper representation will describe the input to the ability in terms closer to what we take to be its *raw* material. In the present case, one would most naturally think of visual properties, and the representation would consist of a pairing of these with species and genus.

To simplify the comparison with the linguistic case, let us partially abstract from force by pretending that all utterances are sayings. One who understands a language must be able, when suitably presented with any particular utterance, *u*, of type *U*, where particular *U*-type utterances are used to say that *p*, to come to know that the speaker who uttered *u* thereby said that *p*. He recognizes the force as *saying* and the content as *that p*. Setting aside force completely (though still only to simplify the formulation), the necessary recognitional ability is that of recognizing the content of utterances. A very shallow representation would take the form: if any L-speaker, S, sufficiently close within his auditory range, uses an L-utterance whose content is that *p*, one with the relevant ability in L is able to know that S uttered something whose content is that *p*. A deeper representation would associate L-utterances, described in non-semantic terms, with their contents. In short, such a representation would be a theory of meaning, though we have as yet found no trace of the requirement that the theory should exhibit semantic structure.<sup>3</sup>

One might hope to derive this requirement from two further premises:

- (P1) The semantic structure of a language is isomorphic with the causal structure of the competence of one who uses it.
- (P2) A full representation of an ability must exhibit its causal structure, which in this case will be the very structure, mentioned in P1, which semantic structure must reflect.

<sup>3</sup> I gloss over the point that my original definition of a theory of meaning made its pairing theorems apply to sentences, whereas this already presupposes discerning structure in the raw data of utterances, any one of which may contain many sentences.

If one can exhibit a causal structure without exhibiting it *as* causal, then the desired conclusion follows fairly straightforwardly. Something like P1 cannot, in my view, be denied, though one might well challenge the rather simple connection which P1 itself claims between semantic structure and the structure of competence.<sup>4</sup> However, P2 is certainly false. It rests on the mistaken view that the causal structure of a recognitional ability is essential to its identity. Unless this were assumed, there is no way in which the obligation to represent an ability can impose the obligation to represent that ability's structure.

Before arguing against this assumption, it is well to see just what the relevant causal structure is supposed to be. In English, it will consist in such facts as these: an English speaker's understanding of the sentence "John runs" causally could not be impaired without thereby impairing his understanding of all such sentences as "Harry runs", "George runs", "Mary runs" . . . , or else of all such sentences as "John walks", "John talks", "John laughs" . . . . A speaker's understanding of "John runs" is thus causally bound up with his understanding of other sentences containing "John" and "runs", and a language's semantic structure is rightly seen as dependent, directly or indirectly, on facts of this kind (cf. Davies 1981b).

I shall first try to show that it is not generally true that the causal structure of a recognitional ability is essential to it, and then I shall try to show this specifically for the case of the linguistic ability under discussion.

Let us suppose that there are chicken sexers who associate young birds with a number from 0 to 1, taken to represent the probability of that bird being male, and that they cannot articulate the features which cue their response. They might share a common ability: over a given range of inputs (these being visual or tactile properties), they assign the same number. One learnt slowly, requiring at least one lesson from an old hand for each type of input, and the sub-abilities which make up the overall ability are causally independent: it is causally possible that the ability to associate any one set of observable properties with a number fail, yet the other property-number associations remain untouched. Another sexer learnt quickly, and extrapolated to associating a number with hitherto unperceived combinations of properties along lines reflected in the theory of probability, and there are causal dependencies among the sub-abilities: there are pairs of sets of properties such that it is causally impossible, as things are with him, that his ability to associate one set of the pair with the appropriate number should disappear yet his ability to associate the other set of the pair with the appropriate number survive. Here we have a clear case in which the same ability is realized in distinct causal structures.

To transpose the analogy in the most immediate way to the case of linguistic ability one must consider a finite language (i.e. one with only finitely

4 P1 would probably have the implausible consequence that something that would normally be considered a single language would not have a unique semantic structure.

many sentences). One speaker has come to learn it on a sentence-by-sentence basis, and his sub-abilities are causally independent. Another has learnt the language on the basis of its structure and there are causal dependencies among his sub-abilities. Yet each might have this ability: suitably exposed to any utterance in the language he can recognize its content.

Extending the case to infinite languages is more controversial, but I can see no logical impossibility in a person's having the ability for an infinite language in a causally structureless way: God might make him know the content of utterances as wholes, yet keep the sub-abilities causally independent. God could, for example, make a person recognize the content of "John runs" and "Harry walks" and not recognize the content of "John walks". If there are logical impossibilities here they permeate the concept of an omnipotent and omniscient being. Fortunately, my case does not turn on such an extreme example as a structureless ability with an infinite language. To convince ourselves that, even in the infinite case, the ability is independent of its causal structure, we can imagine a language with a suffix "-ette" which, attached to any simple predicate "F", forms a complex predicate with the meaning "little F". One speaker's recognition of the content of the (finitely many) "-ette" complexes might be based on a sensitivity to this fact of semantic structure, a sensitivity recorded by such counterfactuals as: if he were to come to fail to recognize the content of "F-ette (this)", while still recognizing the content of "F (this)", he would also, and thereby, come to fail to recognize the content of "G-ette (this)", "H-ette (this)", etc. Another speaker's recognition of the content of utterances containing "-ette" complexes might be insensitive to this aspect of semantic structure: he has had to learn the meaning of each "-ette" complex separately. The common ability of the two speakers extends over infinitely many sentences, but the causal structure of it is different in the two cases.

One who would insist that a theory of meaning exhibit semantic structure might, when faced with this difficulty, simply claim that an exhibition of a structure isomorphic with the structure of competence is of intrinsic interest. This move, however, would seem to sever, or at least make it hard to explain, the connection between theories of meaning and the philosophy of meaning. Let us, rather, consider whether the structural constraint can be derived from the requirement that a theory of meaning should state a condition knowledge of which would suffice for understanding. If understanding a semantically complex sentence required a sensitivity to its semantic structure the derivation would be made. But this dependence of understanding on sensitivity to structure is open to challenge.

Christopher Peacocke has argued for the dependence, deploying the following example (1976: 170). If Brezhnev tells you (in English, for you do not speak Russian) that in the first sentence of his speech (call it "s") he will say that production doubled in 1973 you will know, as the speech begins, what Brezhnev says by uttering *s*, so you will recognize the content of the utterance,

but, according to Peacocke, you will not understand it. The example would indeed establish the failure to understand if we build into it the feature that you cannot identify tokens of *s* as tokens of *s* on the basis of their auditory characteristics, so that, for example, you would be at a loss to say whether Brezhnev uttered *s* just once or more than once in the course of his speech. But if we build in this feature, then the case ceases to be one in which knowledge of what Brezhnev said by uttering *s* amounts to recognition of the content of the sentence- or utterance-*type*, *s*. For it ceases to be the case that *whenever* you are suitably confronted with a token of *s* you recognize the content of the linguistic act the speaker performed by making the utterance. (I do not need to assert that there could not be “one-off” recognitional abilities, independent of the ability to re-identify (“*another death cap*”, “*Margaret Thatcher again*”). I need only assert that the linguistic ability I have discussed is not of this kind.)

What Peacocke could certainly show is that from the fact that you know what Brezhnev said by a token of *s* one cannot deduce that you understood *s*, but this could be established by an example which would make clear that the point does not in itself drive a wedge between a recognitional ability relating to an utterance-*type* and understanding that type: for example, suppose you are told merely that in the first sentence of his speech yesterday (i.e. in uttering *s*) Brezhnev said that production doubled. You plainly could know this without having any access to the auditory or visual properties of *s*, and thus without having the beginnings of the recognitional ability relating to *s*, for this depends on being able to have the relevant knowledge of content on the basis of suitable auditory or visual presentations of *s*.

Suppose someone insists that you could recognize the content of every token of *s*, on the basis of its auditory or visual properties, yet not understand *s* if your recognition was not based on a sensitivity to the semantic structure of *s*. Then, I think, the issue will degenerate into a terminological one about how “understand” is to be used. I have no particular objection to reserving the word for cases in which not only can a person recognize the content of an utterance, *U*, on the basis of its visual or auditory properties, his recognition is, in addition, based on sensitivity to the semantic structure of *U*, a sensitivity manifested by causal dependencies between the understander’s grasp of *U* and his grasp of other utterances with semantic elements in common with *U*. In return, I would have it conceded that one who has the recognitional capacity for every utterance in a language, yet in whole or part lacks sensitivity to semantic structure, is none the less able to use that language for every communicative purpose for which language serves us.

My claim is that a structural constraint on theories of meaning does not follow satisfactorily from either of the theses (*a*) or (*b*) – not satisfactorily from (*b*) because the derivation would involve a somewhat *ad hoc* stipulation about understanding, not justified in terms of the essence of the function of language. But I leave open the possibility that the constraint be derived from

elsewhere, for example from the plausible hypothesis that if we are engaged in the task of radical interpretation it is only by guessing at semantic structure that we can reduce the indeterminacy of interpretation to manageable proportions: to the extent that structure is discerned, evidence for an interpretation of an utterance-type  $U$  will flow not merely from data concerning uses of tokens of  $U$ , but also from uses of tokens of other utterance-types having elements in common with  $U$ .<sup>5</sup> The source of such a constraint will help determine its place in the philosophy of meaning. The source just mentioned would at first glance seem to lie within the epistemology of language rather than its metaphysics and, from a realist perspective on these matters, these two locations are distinct: if there could be a determinate language which *we* could never interpret (to a reasonable degree of determinacy), a constraint flowing merely from the possibility of interpretation will not pertain to the essence of language. I shall note, but not discuss, the delicate issue that now arises: if it is part of the essence of language to be usable for communication, then there can be no absolutely uninterpretable language, but this may perhaps leave logical space for the notion of a language, uninterpretable by us, but interpretable by other sorts of beings (e.g. infinite beings).

## 2 Extensionality

Should a theory of meaning for L avoid non-extensionality (save at most in so far as this is required by non-extensionality in L)?<sup>6</sup>

I stipulate that the extension of a name or singular description is its bearer or denotation, of a predicate the things of which it is true, and of a sentence its truth value. A language is extensional if its every sentence is, and a sentence is extensional if, for each of its parts covered by the stipulation, only the extension of these parts is relevant to the extension of the sentence as a whole.<sup>7</sup> The test is that the replacement of any such part by a coextensive one

5 Gareth Evans (1975) argues convincingly that consideration of the obligation to uncover structure would require an interpreter to discern an ontology of rabbits in a case in which, were the obligation neglected, full Quinean indeterminacy would reign.

6 Evans and McDowell, in their editorial introduction to *Truth and Meaning* (1976), express a preference for an extensional theory, but do not attempt to justify it. John Foster's "Meaning and Truth Theory" contains an argument for the preference which I discuss below. Cf. Platts (1979: 53, 55). Davidson's preference for extensionality in "Truth and Meaning" would appear to be founded more on (unstated) technical grounds than on grounds of increased insight. In this paper I do not consider technical justifications for extensionality, but I know of no good ones. For a non-extensional theory see, e.g. John Wallace (1978).

7 This definition of extensionality is crude. For example, Frege's semantic theory can justly be called extensional, in virtue of its preservation of the principle that only the reference which, in that context, the referring parts of a sentence have, bears on the reference of the sentence as a whole. But no more refined notion is required for the present discussion.

leaves the extension of the sentence as a whole unchanged. It is an immediate consequence of this definition that it is contradictory to hope that a theory of meaning for L should meet all of the following conditions:

- (a) that it state the meaning of the L-sentences;
- (b) that its pairing theorems have the overall logical form “ $s \dots p$ ”;
- (c) that its pairing theorems be extensional.

The reason is that if a sentence “ $s \dots p$ ” states the meaning of  $s$ , its truth value must be sensitive to the meaning of  $p$ , and not merely, as extensionality requires, to the truth value of  $p$ . One who would adhere to extensionality must either abandon the hope that his theory of meaning will state meanings, or else abandon the idea that its pairing theorems will have the logical form “ $s \dots p$ ”.<sup>8</sup>

One reason that has been offered for adhering to extensionality is that mastery of non-extensional idiom “requires an implicit grasp of the . . . concept of meaning” (Foster 1976: 6). The reason would be good only on the assumption, to be challenged in the next section, that a theory of meaning would fail of its purpose if it contained the concept of meaning. But even if we accept this assumption, Foster’s reason is still inadequate. For there exist non-extensional idioms, including one which Foster himself cites, namely “It is necessary that . . .”, mastery of which do not require implicit grasp of the concept of meaning. I can envisage only two ways in which such implicit grasp might seem to be required. On the one hand, it could be held that modal sentences are true wholly in virtue of meanings, and that grasping an idiom requires implicit grasp of what makes sentences containing the idiom true. The shortest way with this suggestion is to consider an idiom of causal modality, which will introduce non-extensionality, but concerning which it would be absurd to suggest that sentences containing it are true wholly in virtue of meanings. On the other hand, it might be held that an idiom requires implicit grasp of the concept of meaning if recognition of the validity of inferences involving the idiom requires sensitivity to difference or sameness of meaning. However, not this sensitivity, but rather sensitivity to the presence or absence of necessary coextensiveness, is required in the case of inferences involving modality. So non-extensionality does not, *per se*, introduce the unwanted concept of meaning.

One might rely on a converse claim to support the extrusion of non-extensional idiom from a theory of meaning: that extensionality at least ensures that the concept of meaning is absent. An initial response might be that this

8 John McDowell takes the first option (1976: 45–7 and esp. n.14). John Foster (1976) takes the second option. A quite different version of the second option is available to one who accepts Davidson’s paratactic analysis, as I point out shortly below.

begs the issue of whether or not a reductive analysis of meaning is possible, for example, an analysis initially in terms of psychological concepts, these giving place to behavioural ones. In the face of this response, one might wish to distinguish the explicit occurrence of the concept, as when it is introduced by some particular short phrase or construction, and its implicit occurrence, as when, as a result of analysis, the concept is, so to speak, dispersed throughout a sentence or set of sentences and is not specifically associated with any short component. The suggestion may then be that all that is objectionable is the explicit occurrence of the concept of meaning, and that this is avoided if the syntax of the theory is extensional.

An *ad hominem* point against those who advance this suggestion yet accept Davidson's paratactic analysis of *saying that* (1969) is that this style of analysis can translate away the non-extensionality of the idiom "means that" without thereby translating away the explicit occurrence of the concept of meaning. Paratactically construed, "*s* means that *p*" will be seen as being (or as being translatable by) two sentences, "*s* means that" and "*p*", where, in the first, "that" is a demonstrative pronoun referring to the subsequent utterance of "*p*". Since the meaning idiom is confined to the first sentence, and since this sentence is extensional, the concept of meaning, on this construal, can occur explicitly yet not be responsible for the introduction of non-extensionality.

### 3 "Means that" and theories of meaning

Both arguments in favour of the extensionality constraint assumed that a theory of meaning for L should make no use of the concept of meaning (except so far as is required by the occurrence of the concept in L), and in particular that the theory's pairing theorems should not have the form "*s* means that *p*", or anything synonymous with this. Davidson speaks in "Truth and Meaning" of the obscurity of "means that" in the course of the discussion which leads to the advocacy of truth theory. Foster more explicitly claims that the occurrence of "means that" would diminish the light we could expect a theory to throw on the notion of meaning (1976: 6). The fear seems to be of circularity. In this section I argue that this fear springs from a faulty conception of what could be hoped for from a theory of meaning,<sup>9</sup> and that one ought to prefer a "means that" theory if one hopes that a theory of meaning

9 The conception is faulty only against the background assumption that either the philosophical account of meaning involves no reduction of that concept, or else the reduction occurs not in the theory of meaning but in the account of what it is for a theory of meaning to be correct. Foster's own position is best seen as one which rejects this assumption, and aims to reduce meaning to intensional structure, within a theory of meaning (cf. Postscript to this chapter). This aim would of course be thwarted by the explicit occurrence of the concept of meaning in the theory of meaning.



should state something, or at least contribute essentially to the statement of something, knowledge of which is sufficient for understanding.

If a theory of meaning for L stated the meaning of each L-sentence, for example by having pairing theorems of the form “*s* means that *p*”, then, if true, it would certainly state something knowledge of which suffices for understanding L. But in the tradition which I am discussing it is generally assumed that the pairing sentences will be T-sentences, having the form “*s* is true  $\leftrightarrow p$ ”, and thus will not state the meanings of L-sentences. Rather, the idea is that by imposing certain constraints on the theory, in each pairing theorem the used sentence in *p*-position will have the same content as the mentioned sentence designated by the term in *s*-position. The T-sentences will thus “fix” the meanings of the sentences they mention, without stating them.

Let us use “D” to abbreviate the predicate expressing the constraints which, supposedly, ensure that all the pairing theorems fix meanings, or, as I shall say, that the theory is interpretative. Then one way in which a non-meaning-stating theory,  $\theta$ , might contribute to a sufficient condition for understanding is as follows: one must know  $\theta$ , and know that D( $\theta$ ).

John Foster (1976: 19) has pointed out a difficulty with this proposal. One might know  $\theta$ , that is know the propositions that  $\theta$  states, and know that D( $\theta$ ), yet not realize that  $\theta$  states these propositions (e.g. if  $\theta$  is in a language one does not understand), and so not know enough for understanding. I shall ignore this difficulty (though this will lead to some strictly nonsensical oscillations between sentences and propositions), assuming that it can somehow be overcome. Instead, I shall press two consequences of any proposal of this general form.

First, such a proposal can be adequate only if it is knowable apriori that for all  $\theta$  if D( $\theta$ ) then  $\theta$  is interpretative. To know a theory (i.e. to know the propositions stated by a theory) is to know (the propositions stated by) its axioms, and to be able to come to know any (proposition stated by a) theorem, by a process wholly apriori relative to knowledge of the axioms. Likewise, a proposal of the present form must state conditions sufficient to enable one to extract the meaning of the L-sentences, by a process wholly apriori relative to knowledge of the stated conditions. Only in this way can the obligation to state a condition knowledge of which would suffice for understanding be discharged. Allowing that knowledge of  $\theta$  will enable one, on confrontation with an arbitrary sentence *s*, to work out  $\theta$ 's pairing theorem for *s*, and that knowledge that D( $\theta$ ) will enable one to work out that this is a pairing theorem of a D theory, one cannot proceed unless one knows that the pairing theorems of D theories fix meanings, and if this is not stated explicitly in the conditions for knowledge, it must be extractable from them apriori if the conditions are sufficient.

In effect, the whole burden of furthering our understanding of meaning in general is borne, in such an account, by whatever light “D” sheds on “interpretative”. But any such elucidation would be just as good an elucidation of

what it is for a meaning-stating theory to be true as of what it is for a non-meaning-stating theory to be interpretative. This holds in virtue of the obvious apriori truth that “ $s \dots p$ ” is fit to be a pairing theorem of an interpretative theory iff “ $s$  means that  $p$ ” is true. Hence a proposal of this kind would licence, rather than prohibit, the use of “means that” as the filling of pairing theorems.

Secondly, not only is this permissible, it is mandatory if knowledge of a theory of meaning is going to be a necessary part of a minimal sufficient condition for understanding. For if we accept any proposal of the kind under discussion, knowledge of  $\theta$  becomes otiose. So long as, in knowing that  $D(\theta)$ , one knows enough to identify  $\theta$ -theorems as  $\theta$ -theorems, one need not know the axioms or theorems of  $\theta$ . The reason is that it is enough for understanding to know that, e.g., the proposition that “snow is white” is true  $\leftrightarrow$  snow is white is a theorem (or rather is expressed by a theorem) of a  $D$  theory. For then one already knows enough to know that this is a meaning-fixing proposition (true or false) and thus that the sentence “snow is white” expresses the proposition that snow is white.

Davidson has to some extent accepted this etiolation of the role of a theory of meaning in providing a sufficient condition for understanding. In “Radical Interpretation” he allows that knowing a theory does not enter into this condition. Rather, he claims, it is enough for understanding a sentence  $s$  that one know that  $s$  is true  $\leftrightarrow p$  “is entailed by some true theory that meets the formal and empirical criteria” (1973: 139); in my terminology, that this is entailed by some true  $D$  theory. But the insistence on the truth of the theory is otiose. For suppose a *false* theory, e.g. one with a theorem saying that “snow is white” is false  $\leftrightarrow$  snow is white, satisfies “ $D$ ”. Then one who knows this can come to know that the sentence “snow is white” expresses the proposition that snow is white.

“ $D$ ” might itself contain the requirement of truth; or we might in any case decree it as a further condition. In the first case, “ $D$ ” would go beyond what an account of interpretativeness requires. In either case, the addition could not be motivated by the search for a sufficient condition of understanding.

An interpretative theory, even if false, can always be transformed into a theory with true pairing theorems, indeed into a theory with T-sentences as theorems. For if the original false, but meaning-fixing, theorem is “ $s \dots p$ ”, it follows that “ $s$  means that  $p$ ” is true, and so is “ $s$  is true  $\leftrightarrow p$ ”.<sup>10</sup> But the fact remains that if only interpretativeness concerns us, given in terms of “ $D$ ”,

10 John McDowell (1976: 46) makes this last inference rest on the claim that “‘ $p$ ’ is true” says the same as “ $p$ ”. But the claim is false, as can be seen from the fact that one could know that  $p$  without knowing that “ $p$ ” is true, and know that “ $p$ ” is true without knowing that  $p$ . Fortunately the inference does not require this falsehood, but only the truth that “‘ $p$ ’ is true (in L)” has the same truth value as “ $p$ ” has (in L).

then the filling of meaning-fixing theorems is irrelevant. A D theory must yield pairing theorems, but neither their overall content nor their truth matters. This affects any non-meaning-stating theorem. So only if a theory is meaning-stating can its truth play a non-redundant role in the formulation of a sufficient condition for understanding. This is one answer to the question why one should hanker after a theory which states meanings.<sup>11</sup>

A corollary is that one for whom a truth theory plays the role characterized by the remark that “truth is what a theory of sense is a theory of” (McDowell 1976: 47) has, in virtue of this alone, no commitment to realism or to the rejection of bivalence. The reason, affording a more immediate route to the conclusion than the one which McDowell himself offers, is that even a verificationist, anti-realist or intuitionist can accept the inference from “*s* means that *p*” to “*s* is true  $\leftrightarrow p$ ”. What he will perhaps reject is a realist or knowledge-transcendent interpretation of the latter.

There is still a positive though slender role for theories of meaning in the philosophy of meaning, as can be brought out by considering an oversimplified version of an account that McDowell (1977: 161) has offered of, in effect, the predicate “D”:

(M) “D( $\theta$ ), relative to L-users” abbreviates “No theory other than  $\theta$  can be used, by treating its pairing theorems as fixing meanings, to give a better account of the behaviour of L-users in terms of their propositional attitudes”.

Compare this with what would result if one tried to apply these ideas directly to the “means that” scheme:

(M’) *s* means that *p* among L-users iff no other interpretation of *s* enables us to give a better account of the behaviour of L-users in terms of their propositional attitudes.

M’ is obviously false, for it would permit adjustment of the interpretation of *s* (e.g. so that it no longer expresses a weird belief of L-users) without a thought for the interpretation of sentences other than *s* sharing some of its syntactic parts. But if M is false, its falsehood is not in this way obvious, at least for structured languages.

From the perspective of one who would advance an account along the lines of M of the enterprise of interpretation, the role of a theory of meaning consists wholly in ensuring that the enterprise relates to language as a structured whole and not to individual utterances considered apart from their structural interrelatedness. (And it will perhaps be said that if there are

<sup>11</sup> McDowell (1977: 161) implies that the hankering is groundless.

structureless languages, their interpretation is, for us, open to an indeterminacy no less radical than that alleged by Quine.) But this role could be adequately fulfilled by a theory whose pairing theorems have the form “*s* means that *p*”; and we have already seen that there is a positive reason for preferring a meaning-stating theory of this kind.

### Postscript: on truth and “truth conditions”

If “truth conditions” is so defined that any truth of the form “*s* is true  $\leftrightarrow p$ ” states the truth conditions of the sentence designated, then no one has ever seriously supposed that a sentence’s truth conditions come at all close to its meaning. If “truth conditions” is so defined that only an interpretative theory yields truth conditions, it is trivial that what yields truth conditions fixes meanings.

These terminological points should not be allowed to obscure substantial issues, notably:

(i) According to the common usage of “truth conditions”, stating a sentence’s truth conditions involves stating a genuinely necessary and sufficient condition for its truth, where such a statement involves modality, and is, perhaps, initially representable as the result of applying the operator “necessarily” to a T-sentence (Foster 1976: 11). It is trivial that, thus understood, the truth conditions of sentences do not amount to their meanings (for all necessary truths have the same truth conditions). But it is a substantial and controversial contention that the way in which a sentence’s truth conditions are fixed by its semantic structure determines its meaning.<sup>12</sup> The issue reduces to that of whether intensional isomorphism is necessarily coextensive with synonymy, which in turn involves the question whether there could be genuinely simple cointensive but non-synonymous expressions.

(ii) One sometimes hears the suggestion that, inextricably bound up with the claim that a Davidsonian truth theory can serve as a theory of meaning is the view that truth can thereby be made to illuminate meaning.<sup>13</sup> Davidson does not advance this view explicitly in “Truth and Meaning”, and the following suggests that he would reject it: “It may be that the success of our venture depends not on the filling but on what it fills” (1967: 23).<sup>14</sup> McDowell explicitly rejects it: “The concept of truth as such . . . need not figure in the

12 Commitment to this contention can inhabit quite different formal semantic frameworks. Compare e.g. Foster (1976: §4) and David Lewis (1972).

13 It is not easy to understand Dummett (1976) except as making this association, though of course he goes on to challenge the views he associates.

14 This is in line with the position of “Radical Interpretation”, which is easily read as placing the whole weight of illumination on the constraints rather than the content of the theory. But it is out of line with the following: “a theory of truth patterned after a Tarski-type truth definition tells us all we need to know about sense” (Davidson 1979: 9).

certification of a theory as a theory of sense” (1976: 146; cf. 1977: 160). This rejection would also follow from the consideration I advanced in the last section, where it was argued that in a theory which failed to state meanings, the filling of the theory, and even its truth, would drop out as irrelevant. One who would try to use truth, within a theory of meaning, to illuminate meaning, would do best to opt for the controversial contention of (i). The concept of meaning would not figure explicitly in a genuine truth conditions theory, yet it is plausible to argue, if the contention is accepted, that knowledge of such a theory would suffice for understanding the language. The contention thus promises an analysis of meaning in terms of truth together with ancillary notions, notably modal ones (which might in turn yield to further analysis; cf. Foster 1976: §4).

It is just possible that by using phrases like “truth conditional semantics” to apply not only to theories of genuine truth conditions but also to Davidsonian truth theories one would slip into the error of supposing that in the latter the concept of truth might reasonably be hoped to illuminate the concept of meaning.<sup>15</sup>

15 This error must be distinguished from the plausible claim that illumination of meaning by truth can be provided by informal accounts. I wish to thank the following for comments on drafts of this paper: Martin Davies, John Foster, Hans Kamp and David Wiggins.

## Evans on reference

---

### I Singular terms

According to Evans, Frege's theory of *Bedeutung* (which Evans translates as "Meaning" or as "semantic value") starts with the idea that the significance of a complete sentence consists in its being true or false (1982: 8).<sup>1</sup> Given this starting-point, Evans continues, it was natural that Frege should think of the semantic power of an expression as the power to affect the truth-value of sentences in which it occurs: and natural, as a further step, to think of this power as determined by the expression's being associated with some extra-linguistic entity. This entity he called the *Bedeutung*. For singular terms, he identified it with what one would ordinarily call their reference, an identification which Evans says "makes evident good sense" (p. 9) provided that we are setting aside non-extensional contexts.

Frege argued that a full account of the significance of an expression must appeal not just to semantic value (*Bedeutung*) but to some further property, which he called *sense*. Sense has to explain "difference in cognitive value" between pairs of sentences, like "Hesperus is Hesperus" and "Hesperus is Phosphorus", which are point-by-point indistinguishable in respect of semantic value. For Evans, the essential notion of Fregean sense is *the way in which the semantic value is presented*. Thus two singular terms which have the same semantic value, and so refer to the same thing, may differ in sense: they may present this referent in different ways. Perhaps "Hesperus" presents its referent in a way connected with evening sightings, whereas "Phosphorus" presents its referent in a way connected with morning sightings. The different modes of presentation may prevent one who has mastered both terms, and so knows for each what its referent is, from realizing the identity.

1 Even if we ought to see this as the fundamental idea in Frege's semantics, it is not presented by Frege in this light. In "On Sense and Reference", Frege starts with *Bedeutung* as a property of singular terms, and only subsequently and derivatively goes on to argue that complete sentences also possess this property.

How should one state what the sense of an expression is? Evans approvingly quotes Dummett: “For Frege, we *say* what the referent of a word is, and thereby *show* what its sense is” (Evans 1982: 26). This idea gives no leverage to those who seek to identify the sense of a singular term with some descriptive condition. In particular, if, following Evans, we take mode of presentation as what is essential to sense, he is surely right to say that there is a contradiction in Frege’s concession, albeit equivocal and qualified (p. 38), that an expression can have a sense yet lack a referent. For if there is no referent, there is no such thing as how the referent is presented.

To understand how Frege was forced into this contradictory position, we must recall two things: that he unhesitatingly grouped together, into a single semantic category, both definite descriptions like “the greatest prime number” and ordinary names like “Aristotle”; and that, in addition to thinking of sense as mode of presentation of semantic value, he also treated possessing sense as equivalent to being significant. It is obvious that there are significant definite descriptions which have no referent. Using the equivalence of sense and significance, this means that definite descriptions can have sense while lacking a referent. And then given the homogeneous treatment of descriptions and names, it would seem inevitable that it should be possible for a name to have sense while lacking a referent.

A common response to this aspect of Frege’s thought has been to cease to take with full seriousness the account of sense as how the semantic value is presented and to see Frege as, rather, identifying the sense of a name with some descriptive condition, thus releasing sense from reference. Evans rejects this response, which sees Frege as a forerunner of Ayer and Searle (p. 27), as wrong-headed, on the grounds that it gives more weight to the rather few passages in which Frege, equivocally and with qualification, concedes the possibility of sense without reference, than it does to the central passages in which sense is said, unequivocally and without qualification, to be how a referent (or, more generally, semantic value) is presented. Evans therefore sees the concession of sense without referent as an aberration on Frege’s part. The remedy involves subdividing Frege’s broad grouping. Although Evans does not put it in quite this way, what one will look for is a group of expressions owing their significance to being non-derivatively associated with a semantic value which is also their referent in the intuitive sense. No doubt an empty definite description, like any other definite description, has a derived semantic value, derived from the values of its parts. But it has no non-derivative semantic value, and obviously its derivative semantic value does not coincide with its referent (in the intuitive sense); so it would be excluded from the category.

Thinking just of singular terms whose semantic value is non-derivative and is simply their referent, Evans’s Frege is in one way close to Russell. For the property of owing significance to association with an extra-linguistic entity, the referent, is just the semantic function characteristic of Russell’s

logically proper names. Russell provides a test for such an expression: to suppose that it lacks a referent is to suppose that it lacks its actual significance. In Russell's words, quoted by Evans:

Whenever the grammatical subject of a proposition can be supposed not to exist without rendering the proposition meaningless, it is plain that the grammatical subject is not a proper name, i.e. not a name directly representing some object. (Evans 1982: 43).

Evans calls a singular term which passes this test "Russellian". A main theme of the book is to establish that a large class of singular terms, including most ordinary proper names and standard uses of demonstratives, are Russellian. If we start with an intuitive grouping of grammatically subject expressions and then remove the Russellian singular terms, there remain at least two sorts of expression. First, those which deserve the title of singular terms or referring expressions, but which are non-Russellian. Evans doubts if there are many examples in English, but suggests that if a name, say "Julius", were introduced by an explicit descriptive stipulation (e.g. "Let us call whoever invented the zip 'Julius'"), it would count as a referring expression (the argument is given on pp. 49–50), but would be non-Russellian, since even if no one had invented the zip "Julius" could still have its actual significance.<sup>2</sup> Secondly, there remain those which are not singular terms or referring expressions at all: thus most definite descriptions are quantifiers (the possibility of exceptions is discussed in Ch. 9.3).

Russell held that two logically proper names (Russellian singular terms) naming the same object have the same meaning:

. . . if one thing has two names [i.e. logically proper names] you make exactly the same assertion whichever of the names you use. (Evans 1982: 43)

In this respect, Russell differs sharply from Frege. Evans sees this as an unwarranted retrograde inference from the Russellian status of singular terms. On Evans's view, two such terms referring to the same object may

2 Is it the case, as Martin Davies suggested to me, that had it not been for this rather peripheral sort of case, Evans could have regained the notion of the "the essence of reference", a title which McDowell tells us (Evans 1982: vi) Evans came to regard as unsatisfactory for a course of lectures? I think not. The variety most stressed in the book is that within the category of Russellian singular terms: the variety of the modes of identification involved in thoughts needed to understand utterances containing them, and the variety of ways in which associated information and modes of identification affect the conditions for successful communication by means of such utterances. See, for example, what Evans has in mind by variety on p. 305.



differ in Fregean sense. Evans aims to steer between two extremes: he rejects Russell's obliteration of the distinction between sense and reference, yet he does not go to the other extreme of allowing sense without reference. Put more positively, his view of singular terms is that they are Russellian, but he allows that co-referring terms may differ in sense.

Frege held that the sense of a sentence, the thought that it expresses, must be common to speaker and hearer in successful communication. Evans denies this. In the case of ordinary proper names, for example, he suggests that it is sufficient for communication that speaker and hearer think of the right object. It is not in addition required that they both think of it in the same way. So the thought a speaker has in uttering a sentence containing a name may diverge from the thought the hearer has in understanding that utterance, even though there is no failure of communication. The point is that successful communication depends upon the appreciation by the hearer of conventionally relevant features of the utterance. In the case of proper names, Evans claims (briefly at p. 40, and as part of a full discussion in Ch. 11) that the conventions link no one way of thinking of the object with the name. Hence there is room for variations among competent speakers and successful communicators, consistently with what the linguistic conventions require.

In this kind of reflection, one can already discover an argument for the Russellian status of those singular terms which are like proper names in the respect just noted. The claim is that the Fregean condition for the communicatively successful use of such an expression is too strong. By what should we replace it? It seems as if what we must require, for a hearer to understand a speaker in his use of such a term, is that both think of its referent. So if there were no such object, the utterance could not be understood, so nothing would have been said, so the term would have no meaning. Hence such terms are Russellian.

Evans envisages this argumentative strategy as early as p. 72. It is amplified in Ch. 9.5 and it runs through the whole of Ch. 11 ("Proper Names"). But despite the interest of this chapter, the strategy is relatively peripheral to the book as a whole. For one thing he holds that it is not usable for every kind of Russellian singular term.<sup>3</sup> For another thing, he holds that there is an independent, and more illuminating, strategy which reaches the same conclusion through considering the nature of thought, in particular, the nature of the thoughts needed in order to understand utterances containing such singular terms. The greater part of the book is devoted to this issue. The upshot is that

3 The exceptions are those for which it is plausible, even if incorrect, to impose a non-object-involving condition for successful communication. Thus for some uses of "That *G* is *F*" which are in fact Russellian, an opponent may be unpersuaded by the envisaged pattern of argument, for he may hold that a sufficient condition for understanding is that the hearer realize that the speaker has said something which is true if and only if exactly one thing is *G* and it is *F*. Cf. p. 136 and pp. 139–41.

for a wide range of singular terms the kinds of thoughts we must have in order to understand sentences containing them are thoughts that would be simply unavailable in the absence of a referent of the term. So if there were no such object, the term could not be understood, that is, it would have no meaning. So it is Russellian.

A large part of the book is concerned to investigate various kinds of thoughts about objects (“particular-thoughts”, as Evans calls them) and to relate them, in various different ways, to conditions for understanding various kinds of singular term. The book is complicated and (to my mind) difficult, so I will suggest some ways of breaking it up into manageable portions.

Part I (Chs 1–3) consists of a historical introduction relating to Frege, Russell, and recent work. The ground covered corresponds roughly to what has been discussed so far in this notice. The last chapter (Ch. 11) is the discussion of proper names, already mentioned. The penultimate chapter (Ch. 10) is a discussion of existence. Although this forms an essential part of the defence of Evans’s Russellian conception of singular terms, the problem being to explain how there could be true negative existential singular sentences, the argument is fairly self-contained. One could therefore read it before reading the rest, or defer reading it until afterwards.<sup>4</sup>

The remaining chapters can be roughly divided into two groups. Chapters 4, 5, and 9 set the general framework within which the main argument for the Russellian status of many singular terms is conducted. Here the two most important things are Russell’s principle (“A subject cannot make a judgement about something unless he knows which object his judgement is about” p. 89) and the idea of information-based thought and its linguistic image, information-invoking singular terms.

As Evans says, Russell’s principle, as just stated, is of no use without some further elaboration of *knowing which*. For on one reading the principle prohibits nothing (in one sense, of course the baffled detective *knows who* he is thinking about when he wonders who the murderer is: the murderer!); while on another it prohibits too much (in one sense, you may not *know who* the man you can see stealing the safe is). The remaining group of chapters, 6–8, can be seen, in part, as providing the needed analysis, applying it to various ways in which one can identify objects in thought (some demonstrative ways in Ch. 6, recognition-based ways in Ch. 8, and the self as a special case in Ch. 7). These chapters are of great interest independently of their place in the larger structure. None the less, one route through the book is this: pause for breath after Part I, get hold of the main argument for Russellian status, based on the nature of thought, and presented in Chs 4, 5, and 9, fill in the details

4 Admittedly, if one were persuaded of the conception yet found the account of existence unsatisfactory, the appropriate response would be not that the conception must, after all, be defective, but rather that one must find an improved account of existence.

with Chs 6–8 (perhaps skipping Ch. 7 on the first reading); and the last two chapters can be picked off when convenient.

My principal aim will be to discuss what I have just referred to as the main argument for Russellian status. In addition, I shall discuss just one other issue: the account of existence sketched in Ch. 10. I thus neglect a great deal of extremely interesting material, notably from Chs 6–8 and 11. It is a matter of particular regret to have no space to discuss Ch. 7 (“The Self”), which is one of the most exciting in the book. It is notable for pushing further the idea, adumbrated in “Things Without the Mind” (Evans 1980), that thinking of oneself involves thinking of oneself as an element of objective reality. This is bound up with Evans’s conception of self-ascription, one upshot of which is that “our customary use of ‘I’ simply spans the gap between the mental and the physical” (p. 236). But despite the attraction of these themes, I decided that readers, and also the book, would be best served if I confined my discussion to the question of Russellian status.

## **2 Information, communication, thought**

### **2.1 Information-invoking singular terms and information-based thought**

Evans’s account of thought about objects is governed by Russell’s principle:

A subject cannot make a judgement about something unless he knows which object his judgement is about (p. 89).

He holds that there are three central kinds of *knowing which*, or, as he calls them, kinds of modes of identification: descriptive, demonstrative, and recognition-based.<sup>5</sup> These always enter into the determination of the identity of thoughts, though they do not necessarily enter (in any straightforward way: see 2.2 below) into the determination of the identity of the unit of communication.

There is no simple link between singular terms and kinds of mode of identification. Thus an utterance of “That *G* was *F*” may express a thought involving the descriptive kind of mode of identification, if the identification exploits the fact that the object is, or was, uniquely *G* (cf. p. 135). (Note that the descriptive kind of mode of identification must be firmly dissociated from definite descriptions, assuming that these are treated quantificationally. A quantificational phrase, even when it includes a uniqueness condition,

<sup>5</sup> I shall reserve “mode of identification” for something highly specific. But I shall allow “kind of mode of identification” to include more specific kinds than the three just mentioned, e.g., the perceptual kind (a species of the demonstrative kind).

imposes no requirement of identification.) An utterance of the same sentence may also express a thought in which recognition-based identification is involved, for example if I recognize a currently perceived *G* as one previously encountered. And an utterance of the same sentence may, if used in connection with a *G* currently perceived, and not recognitionally identified, express a thought involving demonstrative identification. Where an object is identified by the mode of identification involved in a thought, let us call it the thought's *Object*. Evans uses "object" in this quasi-technical sense (pp. 138–9), though he does not stress it by the initial capital.<sup>6</sup>

Very often, in thinking about an object a subject possesses information flowing from that object. Information includes misinformation (p. 121). A central example is perception, but memory and testimony also provide such an informational link. Information can be supposed to be expressible by an open sentence (pp. 122, 125). Thus the informational content of a photograph of a red ball on top of a yellow square can be reported by the open sentence:

Red(*x*) & Ball(*x*) & Yellow(*y*) & Square(*y*) & On Top Of(*x*,*y*) (p. 125).

Information (with content *Fx*) "controls" a thought of a particular object if and only if the subject, thanks to acquiring and retaining the information, is disposed to "appreciate and evaluate" thoughts about the object as thoughts about an *F* thing (p. 122). For example, suppose a subject is looking at a black and white cat he has not previously encountered or heard about. The content of his perception, and no other information, will control his thought about the cat:

He will be able to entertain the thought that the cat is ginger, but he will, of course, grasp it as false; whereas the thought or speculation that the cat is a favourite of Queen Elizabeth will be grasped as having the (probable) consequence that Queen Elizabeth likes black and white cats; and so on (p. 121).

I presume that informational content could be expressed as a generalization as well as by an open sentence. If so, then we could say that a perceptual informational content, expressible as a perceptual content as of *a* black and white cat, interacts with a speculation (that that cat is a favourite of Queen Elizabeth) to yield a thought recognized as a probable consequence of the speculation. The information, *Fx* or  $\exists xFx$ , is thus linked to the subject's thought of an object in that the subject reasons as if he had the premiss

6 The reference here is to an appendix. Though John McDowell, the editor, warns (p. vi) that some of the material in Appendices may not be verbatim Evans, I treat the book throughout as the work of a single hand.

(which he need not hold explicitly), concerning the object, that it is *F*. (Evans says that we must not hold in general that information possessed is information believed (p. 123). But he holds that for information to control thought it must be believed: “the thought-episode will manifest a belief about the world”, p. 122; cf. also Ch. 9.4 and esp. pp. 331–2.)

In having particular-thoughts which are controlled by information, the subject’s overriding purpose is to be thinking of the object from which the information derives (p. 138). He aims at this object, which Evans calls his *target*, but, as with literal aiming, he may miss it: “as happens when the mode of identification employed does not identify that object” (p. 138). We can now characterize a necessary condition for the existence of an information-based thought about a particular object: in such thinking, Object and target must coincide. That is, the same thing must be that which is identified by the mode of identification involved, and that which is the source of the information controlling the thought. (This simple picture requires modification in the light of Evans’s discussion on pp. 316–18, but I shall ignore the complicating factors.) As Evans puts it:

Only if these two procedures [viz: that which determines Object and that which determines target] locate the same object can the subject be credited with an information-based particular-thought about that object (p. 139).

No such condition governs the having of purely descriptive (i.e. quantificational) thoughts. For the notion of target plays no role in stating the condition for having such thoughts.

Information-based thoughts have a property analogous to the property of being Russellian, as applied to singular terms. If a mode of identification in fact identifies nothing, then it cannot figure in an information-based thought about an object, for the condition of coincidence of target and Object cannot be satisfied. So for an information-based thought having *o* as its Object, had *o* not existed, that thought could not have existed.<sup>7</sup> Exploiting this analogy, Evans calls thoughts meeting this condition Russellian (cf. pp. 71–2). Evans

7 There is room for doubt about the correctness of this interpretation, which sees the Russellian status of information-based thoughts as an immediate consequence of their definition. Evans’s discussion in Ch. 9, and especially 9.4 and the remarks at pp. 326–7, suggest that Evans takes their Russellian status to be a substantive thesis. I suspect that there are alternative presentations. On my interpretation, the substantive issue is whether information-invoking singular terms, characterized in some superficial way, require information-based thoughts for their understanding. On the alternative interpretation, the substantive issue is whether information-based thoughts are Russellian (information-invoking singular terms can be characterized as those requiring information-based thoughts for their understanding). Essentially the same arguments will address either of these substantive issues.

has earlier warned that it is no part of the proposal that one who tries to think such a thought, but fails through lack of an appropriate object, has a wholly vacant mind (p. 45). On the contrary, it may be for him as if he had such a thought, and this may involve images and words passing through his mind and the having of *other* thoughts, for example general thoughts. This view about thought is perfectly consistent with the following view about thought-expression: where an attempt to *express* a Russellian thought (e.g., in the words “This *G* is *F*”) fails because there is no appropriate object, no thought correctly attributable to the subject, on the strength of his utterance, counts as the thought expressed, in virtue of the linguistic conventions, by those words thus used.

So much for the account of what information-based thought is. The role it plays in the main argument is as follows. Evans claims that for many singular terms one must think an information-based thought in order to understand utterances containing them (see esp. Ch. 9.4). Information-based thoughts are, as we have seen, Russellian (and this follows more or less immediately from their definition). Hence the singular terms are Russellian: if they did not refer, there could be no thought sufficient for understanding utterances containing them, that is, they would not be significant. In giving his examples of information-based thoughts in the detailed discussions of Chs 6–8, Evans has, of course, used sentences to express these thoughts. So to some extent the work has been done. None the less, Ch. 9 contains the official general argument.

As expressed in that chapter, the task is to show that “information-invoking singular terms” require information-based thoughts for their understanding. But how are we to identify an information-invoking singular term? At the outset, it should not be so classified in virtue of requiring an information-based thought for its understanding, on pain of trivializing the thesis. Later, of course, we can use the need for information-based thought to effect a principled classification of singular terms. As an example of this general strategy, consider this passage from the discussion of demonstratives in Ch. 6.

If our conception of demonstrative identification is tied in a simple way to the use of demonstrative expressions in English, then we shall include in that category cases like the one in which a man is indicated by means of his footprint. . . . Nothing general may be said about such a rag-bag category (p. 199).

The way to end up with a category with genuine unity is to distinguish it *semantically*. Reverting to our discussion of Ch. 9, however, we need some initial specification of an information-invoking singular term which leaves a substantive thesis to be established.

Although Evans is not very explicit, we can initially say that an information-invoking singular term is one typically intended, as used in an utterance,

to make the hearer bring to bear, in understanding the utterance, information antecedently in his possession. It is far from immediate that such utterances require information-based thoughts for their understanding. For, first, it is one thing for the speaker to intend that information be brought to bear in understanding, and another thing for the bringing to bear of this information to be required for understanding. Secondly, since information includes misinformation it may seem that bringing information to bear does not even require that the hearer have a belief corresponding to the information, and if he needs no such belief he may not need an information-based thought. Thirdly, even if the hearer believes the information, it may seem that it would be possible for the belief to be false, in which case the route to Russellian status will be blocked. On my reading of Ch. 9, Evans addresses the first question by means of examples, and the second two by general considerations.

On the first point, consider the following passage:

If a speaker utters the sentence “This man is *F*”, making a demonstrative reference to a man in the environment he shares with the hearer, the hearer can understand the remark only if he perceives the man concerned, and, bringing his perceptual information to bear upon his interpretation of the remark, judges “This man is *F*: that’s what the speaker is saying” (p. 305).

Although one cannot call this an argument, it strikes me as convincing. At another point, Evans envisages the following situation.

*S* and *A* were in the habit of going hunting together in their youth. On one of their hunting trips, they saw a dazzlingly beautiful bird perched in a pine tree. Years later, *S* (the speaker) may advert to this incident, and say something like: “Do you remember that bird we saw years ago? I wonder whether it was shot” (p. 308).

Evans claims that *A* does not fully understand the speculation (that that bird was shot) until he has connected *S*’s use of the phrase “that bird” with information *already in his possession* (my emphasis), information which relates to that bird. No matter how fully *S* may go on to describe the bird, *A* will not understand the original remark until the connection has been made, a connection often manifested in a flash of recollection.

I think we can take it that these examples show that there are many cases in which understanding requires activation of antecedently possessed information. The examples also suggest that an information link to the referent is required, and this would ensure Russellian status for the singular term. However, we do not have to rely on the examples to establish this.

Given that information must somehow be brought to bear in understanding

an utterance, must a Russellian thought be involved? One alternative suggestion is this. Suppose that in the utterance “*t* is *F*” the use of “*t*” invokes information or misinformation representable as “ $\phi_1, \dots, \phi_n$ ”. Then understanding the utterance could be held to consist in the realization that what the speaker said is true if and only if something is  $\phi_1$  and  $\dots \phi_n$  and *F*. This exploits the invoked information, but does not require, on the part of the understander, a Russellian thought. For, in particular, no commitment is extracted from the hearer that “*t*” refers or that anything answers to the information (cf. p. 328).

Evans allows (p. 329) that this sort of picture is approximately right for a singular term like “Julius”, introduced by the explicit stipulation “Let us call whoever invented the zip ‘Julius’”. The reason is that understanding such a remark requires being faithful to the speaker’s conception of the object, because the speaker’s intention, in using such a term, is that the referent be identified *as* the possessor of the property of inventing the zip. The speaker’s overriding intention is to convey his conception, and this can be achieved even in the absence of any object it concerns (cf. pp. 328–9). But this is not the case with information-invoking singular terms. Though the speaker intends his hearer to bring information to bear, there may be no information concerning which the speaker has this intention, and the information the hearer quite properly brings to bear, honouring the speaker’s intentions, may not figure in the content of a belief of the speaker’s about the referent.

So in bringing to bear the information, the hearer is not merely following a path laid out for him by the speaker. He must draw upon his own resources in order to select *appropriate* information. Evans suggests that this selection must be made as follows:

the only possible justification of the belief that, if what the speaker said is true, there is something which is  $\phi_1, \dots, \phi_n$  and *F* is that it follows from some belief of the form “The speaker is referring to *a*”, together with a view as to how things stand with *a* (p. 329).

In other words, the hearer knows which information to bring to bear in virtue of knowing what the speaker is referring to. He then brings to bear the information he believes pertains to that object.

This is one natural interpretation of Evans. It may appear hopelessly circular. For are we not using the notion of reference to explain the notion of reference? In more detail, was not associated information to be used to explain what is involved in understanding a referring expression? And in this case must it not be incorrect to appeal to a prior identification of the term’s referent en route to an identification of the appropriate information?

I am not sure. It should be remembered that, for Evans, there are many kinds of referring expression. There are non-Russellian non-information-invoking ones like “Julius” and Russellian but non-information-invoking



ones like “here”. The idea of a referring expression in general is anchored by the argument that purported to show that “Julius” should be so classified (pp. 49–50), so there is no general reason why it should not appear in the explicans at this point. And if the more specific version of the worry about circularity is meant to suggest that a speaker could first tell what information was appropriate and then go on to identify the referent, it appears to express a doomed hope: the hope is seen as doomed once it is accepted that understanding a remark may require the bringing to bear of information even though the speaker has no intentions concerning this information, and so cannot manifest in his utterance what information is to be brought to bear. So I think that the fear of circularity is misplaced.

Accepting this much by no means completes the argument. For one objection is: is it not enough that the hearer believes that the speaker is referring to something? He can bring information to bear on the basis of this belief, even if the belief is not true.

Against this objection, Evans argues that understanding is a species of knowledge, and knowledge cannot be based on false belief.

Truth is seamless; there can be no truth which it requires acceptance of a falsehood to appreciate (p. 331).\*

There remains a category of objections of which one example is this. Suppose that it is rumoured that there is someone called “Jack” who is  $F_1, \dots$  and  $F_n$ . Suppose I know part of this rumour and know that it is not merely false, but a total fabrication, not originating in anyone’s actual deeds. Suppose I also know that you know part of the rumour and believe that it is true. If you now utter the words “Jack is  $G$ ” surely I can understand what you say. I know that I am to bring information to bear, so the example is consistent with “Jack” being an information-invoking singular term (on a superficial characterization of this category), and I bring appropriate information to bear. But there is no object involved, and I do not even believe that there is. (This objection thus respects the “seamlessness of truth”, though there are obvious variants which do not.)

Objections of this category Evans classifies as ones involving pretence. He would describe the case not as one in which I understand your remark, but as one in which I pretend to do so. In general, he allows that something *very like* understanding empty information-invoking singular terms (or rather apparent singular terms) can occur in cases of pretence (cf. pp. 330, 339–40). To modify the “Jack” example so that no one is deceived, imagine that you and I share a magic mushroom and are victims of similar perceptual experiences.

\* [Added note] This argument is discussed further in essay IX below.

As hardened users, we recognize these to be hallucinations. If I utter “That little green man is *F*”, it *seems* that the fact that the subject expression does not refer is no barrier to your understanding what I said, and that the use of the expression invokes the information (misinformation, of course) contained in the hallucination. But Evans asks that these cases be set aside for special treatment (Ch. 10; and see §3 of this chapter):

A proper appreciation of the nature and extent of linguistic pretence holds the key to an adequate theory of reference. For it, and it alone, enables us to hold on to the insight that singular terms are (generally speaking) Russellian, while taking a realistic and credible view of phenomena apparently inconsistent with that insight (p. 340).

This is a powerful weapon: it certainly seems that the apparently most compelling counterexamples to Evans’s theory can be brought within this category.

I defer all criticism of Evans’s account until 2.3 below. I now simply note its consequences for communication. This will explain how the view that two co-referring expressions may differ in sense is implemented.

## 2.2 Communication

Frege held that communication requires speaker and hearer to associate the same thought with the words used. Evans departs from this view (see e.g. pp. 315–16). Counterexamples, for him, are proper names, pronouns like “I” and “you”, and even demonstratives like “that ship”, used of a ship before the eyes.

We have two distinct continua. On the one hand, there is one set by the degree to which understanding a singular utterance involves not merely identifying the object, but doing so in a particular way. Many proper names rank low on this continuum: it is not merely that speaker and hearer may communicate successfully while exploiting quite different modes of identification of the name’s bearer, but rather that, for the speaker, almost any old mode of identification is one whose use by the hearer will not thwart the speaker’s communicative intentions. (Not all proper names, or not quite any old mode, else there will be no room for an account of a *misunderstanding* of “Hesperus” as though it meant what “Phosphorus” means.) By contrast, “I” and “that *G*” (used of an object perceived) rank high on the continuum. Understanding an “I” utterance requires identifying the utterer and realizing that he is the referent of the pronoun. One has not understood the utterance if, for example, one comes to think of the utterer, but does not think of him as the utterer. In the case of “that *G*”, understanding requires demonstrative, in the envisaged case perceptual, identification of the referent.

The other continuum is set by the degree to which the mode of identification the speaker himself uses must resemble that which an understander is to use. Proper names are again at the low end of this continuum, for the reasons already given; and “that *G*”, understood as before, is again at the high end. But “*I*”, by contrast, is at the lowest end of this continuum. Where speaker and hearer are distinct, each must use very different modes of identification, indeed different kinds of modes of identification: the speaker uses self-identification, the hearer does not.

Consider two cases. (1) A speaker looking at a ship through one window,  $w_1$ , says “that ship was built in Japan”. His hearer, thanks to his position in the room, cannot see the ship through  $w_1$ , but can see it through another window,  $w_2$ . Their ways of identifying the ship are similar, but not the same. They exploit the same kind of mode of identification, viz., the perceptual kind, but different modes within this kind. However, the similarity is good enough for successful communication: the hearer can be credited with understanding the speaker. (2) A thinker, looking through  $w_1$ , believes something expressible by “that ship was built in Japan”. Looking through  $w_2$  a moment later, he does not believe something expressible by “that ship was built in Japan”. He does not realize that the ship he sees through  $w_1$  is the same as the ship he sees through  $w_2$ . If we are to avoid the contradiction that the thinker both is and is not related to a single thought by the belief relation, we must distinguish the  $w_1$ -thought from the  $w_2$ -thought (cf. p. 84).<sup>8</sup> For Evans, the distinction emerges from the differences in the specific modes of identification. They are similar, sufficiently similar to have permitted understanding of a related utterance, as in (1), but none the less they differ in that on the one occasion the object is located at one place relative to the subject, on the other at another.\*

For Evans, Frege ran together two ideas under the label “thought expressed by a sentence”. One is the idea of what is conventionally required to understand a sentence. Another is what is in the subject’s mind in virtue of thinking something expressible by the sentence. The first idea gives rise to a relatively coarse-grained specification (the degree of coarseness varying from one singular term to another), apt for giving a realistic account of communication (cf. p. 40). The second idea gives rise to a much finer-grained specification, too fine-grained to give a sensible account of communication, but required for a fully specific account of how things are with a thinker.

8 Evans notes that this example is similar to one used by John Perry (1977: 483). It is an accident that the case turns on spatial difference. It is useful to reconstruct it in terms of temporal difference. Evans would presumably have to insist that for the thoughts to differ there will have to be a phenomenal difference with respect to time like that with respect to space.

\* [Added note] Essay VIII below contains further discussion of this example.

Accordingly, one must give a divided answer to the question how Evans implements his adherence to the distinction between sense and reference. If sense is regarded as determining the unit of communication, then the distinction will manifest itself (as perhaps in the case of “Hesperus” and “Phosphorus”) in the conventional association with different modes of identification of two terms for the same referent.<sup>9</sup> There is a more important implementation of the sense/reference distinction, more naturally stated at the level of thought than at the level of language. If thoughts ascribe the same property to the same object, they can none the less be distinct if they involve different ways of presenting the object, that is, different modes of identification.

### 2.3 Two objections to Evans’s view of thoughts

I wish to consider two rather general objections, both discussed by Evans, which serve to make the broad outline of his position stand out more boldly.

As I have said, his view lies between two extremes, and this goes as much for his view of thoughts as for his view of singular terms. At one extreme is Russell’s view, on which the existence of Russellian thoughts is, of course, granted, but it is denied that there can be *two* such thoughts predicating the same property of the same object. At the other extreme is the view which some, though not Evans, attribute to Frege, on which there is no problem about allowing thoughts to be distinct, even though they predicate the same property of the same object, but it is denied that thoughts are Russellian. Let us call these, respectively, the Russellian and the Fregean poles. Evans, of course, has a view combining elements from each pole: Russellian status together with allowing the Fregean distinction.

At the Russellian pole, thoughts about objects are seen as having the identity conditions of sequences: a thought in which an  $n$ -ary property is ascribed to  $n$  objects can be regarded as a sequence of  $n + 1$  entities, the property followed by  $n$  objects. The view must be coupled with an account of the conditions under which a subject can think of an object. One such account, which Evans takes as a prime target, is what he calls the Photograph Model. Evans attributes its popularity to a misunderstanding of Kripke (Ch. 3.4).

It is sufficient for a photograph to be of an object,  $X$ , that  $X$  played a suitable causal role in its origin (cf. p. 78). In a similar vein, Kripke (1972)

9 We need here an intermediate level of classification of modes of identification: we need something more specific than kind of mode of identification, but less specific than a single mode: *ranges* of modes of identification, perhaps.

One must beware of thinking that it follows that there is any purely descriptive specification of such a range of modes. On the contrary, one may be able only to *show* the range, e.g., by re-using the term in a sentence like “‘Hesperus’ refers to Hesperus”. Cf. p. 26 and also McDowell (1977).

suggested that it is sufficient for someone's utterance to involve reference to an object, *X*, that *X* played a suitable causal role in the origin of the utterance. Evans does not enter any objection to this, despite the fact that it is a theory of sayings which fails to enter the kinds of restrictions that would be required by Russell's principle *if* referring to *X* entailed thinking of *X*. For Evans denies the entailment. He claims (p. 74) that if one uses a proper name, facts concerning other people in the name-using community as a whole may make it the case that one has spoken of the name's referent, even though individual facts concerning just oneself (the only facts relevant to what one is thinking) do not ensure that one has really thought of its referent. In general:

The abandonment of [Russell's principle] at the level of saying is a trivial consequence of the distinction between what one says and what thought one intends to express (p. 76, n. 18).

What Evans does object to, however, is that "later synthesisers and bandwagon performers" (p. 79) have confused Kripke's Photograph Model of sayings with a Photograph Model of thoughts. The latter is the view that it is sufficient for someone's thought to concern an object, *X*, that *X* played a suitable causal role in the ancestry of the thought-episode. The view is objectionable, according to Evans, because it abandons Russell's principle and thereby "encourages" (p. 82: he does not say "entails") the sequence theory of thought. This Photograph Model of thoughts is, then, the first of the two general objections I wish to consider.

Evans uses various examples to try to persuade us of the incorrectness of the Model. One is this:

On a certain day in the past, a subject briefly observed two indistinguishable steel balls suspended from the same point and rotating about it (p. 90. For other examples, see pp. 149–59 (hearing on the radio), and pp. 164–7 (the submarine)).

The subject knows no facts which would discriminate the two balls. Can he think of one ball, rather than the other? It is not clear that the Photograph Model need deliver an affirmative answer, since it may be held that in this case there has been a merging of causal lines (whereas what is required, it may be said, for a thought-episode to concern an object, *X*, is that there be a single causal line running from *X* to the episode). However, Evans considers a variant of the case, which does not raise this problem. We are to suppose that the subject saw each ball on a different day, but that selective amnesia has obliterated the memory of one incident. This makes it plain that the origin of the current thought (if there is one) lies in just one of the two balls. But can the subject really think of this ball? The problem, according to Evans, is that:

There is no question of his recognizing the ball; and there is nothing else he can do which will show that his thought is really about one of the two balls (about *that ball*), rather than about the other. The supposed thought – the supposed surplus over the *ex hypothesi* non-individuating descriptive thought – is apparently not connected to anything (p. 115).

By this stage in the argument, Evans has introduced the notion of the “fundamental level of thought” (Ch. 4.4), an account, roughly, of a canonical way of thinking, in terms of which he hopes to elucidate the notion of *knowing which*. The fundamental level presupposes Russell’s principle, so it cannot be used, in any direct way, as a basis on which to reject an alleged counterexample to that principle. We also have in play the “Generality Constraint” (cf. p. 75, n. 15), which requires that if someone can think about an object that it is *F*, he must also be able to think about it that it is *G*, for any concept *G* which he possesses. But simply so far as intuitions go, even as expanded by the generality constraint, there is nothing to be extracted from the case which would serve Evans’s purpose. This becomes plain if we hold before our minds the distinction between whether the subject is really thinking about that ball and whether *he knows that* it is that ball he is thinking of. The Photograph Model claims only that the subject can think of that ball, not that he can know that he is thinking of it. For those already persuaded of Russell’s principle, this will be a distinction without a difference. But the position is meant to be that the adherent of the Photograph Model proposes the case as a counterexample to Russell’s principle. So Evans cannot defend himself by simply drawing upon the principle.

The Photograph Model theorist might argue as follows. If you have seen just one ball, and there are no further relevant circumstances, there can be no doubt but that you can subsequently think of it. Further, if you have seen just one ball, and you also very nearly, though not quite, saw a second indistinguishable one (you would have seen it had not someone quite fortuitously decided to put it into a cupboard), then again you must be able to think of the only one you saw. A mere possibility cannot affect the actuality of thought. But how does this differ from the case under discussion, in which though two balls are seen, the memory of one incident is obliterated? Since the second ball now impinges in no way upon your consciousness, its nullified impact can make no difference to whether or not you can think of the first one.

The problem for Evans is this: can he allow the uncontroversial fact mentioned at the beginning of the previous paragraph without thereby committing himself to the rest? If the subject encounters just one ball, the fact remains that he has at best a seemingly low-grade recognitional capacity with respect to it: he could not distinguish it from other indistinguishable balls. Matters are made no *worse* by changing to the case in which the subject has encountered two balls: with respect to either ball, he has only the low-grade recognitional capacity he had when only one ball was involved. The

upshot is that, for Evans, what makes the difference between the two cases – the first being one in which the subject can think of the ball, the second being one in which, according to Evans, he cannot think of the only ball that is currently impinging upon him – cannot flow from any qualitative difference in the recognitional capacities.

Evans grapples with what is in effect this problem in Ch. 8.3 and also in the Appendix to Ch. 8. The solution of the Appendix is to say that more is involved, in the capacity to think of an object in this sort of case, than the (defeasible) ability to know, if one were to re-encounter it, that one has done so. What is additionally required is that one can more or less definitely locate an encounter with it in one's own past (p. 300). The trouble is that this addition does not help in the case of the steel balls (where two balls have been encountered, but the memory of one incident obliterated), since the subject may be able to locate his encounter with the one ball which now impinges upon him with an accuracy which, while insufficient to distinguish the encounter from the forgotten one with the other ball, satisfies any reasonable general condition. Suppose, for example, that the incidents were a few moments apart. Any reasonable general condition is met by one who correctly holds that the encounter occurred when he was visiting the metallurgy department in London on Wednesday last between 11 and 12 in the morning.

It may be that Evans was wrong to think that his theory requires him to diverge, in this case, from the Photograph Model. The essence of the position he has to defend, against the Model, is that having a particular-thought is knowledge-involving: you must know which object your thought concerns. This contrasts with the view of the Model, according to which having a thought is like perceiving: it may take place even when you are in no position to gain knowledge of the object involved. Evans must therefore hold that a situation which would prevent any knowledge of an object would prevent any thought of that object. But it is unclear that this is threatened by the case of the steel balls. We might well agree that the earlier version of the case, in which there is no amnesia, is one in which the possibility of knowledge is destroyed. But we have already seen that the case is not then one in which it is clear that the Model entails that the subject can think of one of the balls. And we might feel as disinclined to think that the second version, with amnesia, is knowledge-destroying as we would be to hold that a man who cannot tell fake ducks from real ones, and who is in an environment where there were once plenty of fakes, but these have now all been destroyed, cannot now know, upon encountering a (real) duck, that he is encountering a duck. More generally, disagreement here may be not over the question of whether particular-thoughts are knowledge-involving, but rather over the conditions under which knowledge is possible.

Evans agrees that not much should hang on unprincipled intuition, especially as, he claims, we tend to confuse the phenomenon we should be

concerned with (whether a subject can think of an object) with unrelated questions, like whether he has a certain object “in mind” (see Ch. 5.3).

The Photograph Model is said to “encourage” the sequence theory of thought, and Evans has a decisive refutation of this theory. We saw above that it is quite coherent to suppose that a subject is belief-related to a thought ascribing a property to an object and not belief-related to a thought ascribing the same property to the same object. On the sequence theory, the thoughts are the same, and so the supposition would be contradictory, which it is not. Hence the sequence theory is false.

We have here a refutation of the Photograph Model only if the Model entails the sequence theory. Evans gives no elaboration of the Model which would enable us to tell whether this entailment holds. In connection with the kind of case which refutes the sequence theory, Evans says, speaking of the Model, that it is “quite obscure” *how*

the sheer difference between the causal relations could generate a difference in *content* between the two mental states, given that it need not in any way impinge on the subject’s awareness (p. 83).

But we could well imagine the Model augmented in various ways, ways suggested by the original photographic analogy. This spoke of a “*suitable* causal role”. Whatever its causal origin, a uniformly black print is not a photograph of a magnolia in full flower. The “*suitable*” causal role must involve a requirement that information is, to some degree, preserved. This suggests that the Model does not entail the sequence theory. Though not every difference in the causal origin of a thought will “impinge on the subject’s awareness”, the Photograph Model theorist will seize upon differences that do thus impinge. So the knock-down argument against the sequence theory leaves the Model untouched.

What one would need to consider, to decide this issue, is whether a full elaboration of the Photograph Model’s implicit requirement of some level of match between object and consciousness could stop short of requiring Russell’s principle. I incline to think that it can, just as perception can require, in addition to the mere fact of causation, some level of match, without thereby issuing in an account which guarantees that a subject who can perceive an object is, merely thereby, in a position to gain knowledge of it.

I conclude that Evans’s discussion of the steel balls, taken by itself, ought not to cause the Photograph Model to be abandoned. The crucial issue of whether particular-thought is knowledge-involving *could* not be settled by this case, except in the presence of a quite detailed theory of knowledge. Evans’s main contribution to the debate lies not so much in his discussion of this example, but in his detailed discussions of the various kinds of modes of identification.

In reacting against the Russellian pole, represented by the sequence theory



and its alleged ally, the Photograph Model, there is a natural tendency to swing at once to the Fregean pole. What was wrong with the sequence theory was that it left no room for the different ways in which an object can be represented in our thought. But if a thought is to consist in *representations* of objects and properties, then it seems that it cannot be Russellian: for a representation could exist even if it fails to represent anything.

This is precisely the move which Evans rejects. His claim that some thoughts are Russellian is just the claim (transposed into the present terminology) that some representations are so related to their objects that, were the object not to exist, nor could that representation. In “representation of *a*”, “*a*” figures, for Evans, not just as a specifier of causal origin but as a specifier of content.

We . . . need . . . a distinction between, on the one hand, an *a*-representation (i.e. a species of particular-representation, in a specification of whose content mention of *a* would figure: something which represents, or misrepresents, *a*), and, on the other, something which, without being an *a*-representation, is a representation of *a* (p. 125, n. 10).

A view of thought and representation which apparently rivals Evans’s is this. When we think of a particular object, the thought in itself is non-Russellian: it is a thought which, in other circumstances, would have been a thought of a different object, and in yet other circumstances would have been a thought of nothing at all. What makes a thought concern one object rather than another, or concern any object at all, is the causal relationship between the thinker and his environment.

This source of opposition to Evans’s theory could press into service the Photograph Model discussed earlier. The thought will correspond to the informational content of the photograph, where this is conceived to be specifiable quite independently of what objects there are (cf. p. 125). Whether a photograph is of any object at all (as opposed to being a result of darkroom trickery), and if so of what, is a matter of the photograph’s causal history.

There is an analogue with perception. We tend to hold that perceptual experience may have a content irrespective of whether the experience is of an object, or of which object it is of. Facts of the last sort are determined by the experience’s causal history (though, as mentioned earlier, this may be a complicated story, involving such notions as preservation of information).

On the view in question, the supposition that Descartes’s evil genius really exists entails that all our beliefs about the material world are false. On Evans’s view, the particular-beliefs we actually have would be unavailable under this supposition, and so would not be available even to be false.

A theoretical reason sometimes advanced for the anti-Evans view is this: thoughts explain actions, and adequate explanations must draw on only the intrinsic properties of the things involved in the facts to be explained. But

Evans's view fails to respect this requirement, since he holds that particular-thoughts are not intrinsic to the thinker (the existence of the thought requires the existence of something other than the thinker).

This kind of anti-Evans view may be implemented in two ways. It may be held that all thought is really general. Which object a speaker is thinking of, though it is a question which can be made intelligible, has no bearing either upon the identity of the thought or upon its truth conditions. (This is like Russell's view of descriptive thought: in such thinking an object may be "denoted", but which object, if any, is denoted is irrelevant to the existence or truth-conditions of the thought.) There are plenty of arguments against this view, in Evans and elsewhere. One main line of objection is that in many cases it seems impossible to find a general proposition which both has appropriate truth conditions and is also one that it is psychologically plausible to attribute to the subject.

I shall discuss the other way in which the anti-Evans view can be implemented. On this way, it is accepted that no general proposition captures the truth conditions associated with successful particular-thoughts: the relevant truth conditions are irreducibly singular. But it is held that the thoughts themselves owe neither their identity nor their existence to any truth conditions which may be associated with them. Whether a particular-thought has truth conditions and, if so, what these are is a matter of its causal origin. If such a thought has the extrinsic feature of appropriate causal origin involving an object, *o*, then *o* will enter into its truth conditions; otherwise, the thought has no truth conditions.

Evans identifies this view as the dual component theory (cf. p. 201). His main argument against it is that it makes thoughts into schemata. But schemata cannot represent the world as being thus-and-so; for representing is correct or incorrect, and schemata can be neither.

The objection is simple. It is of the essence of a representational state that it be capable of assessment as true or false. If a state is a representational state, it represents something other than itself as being thus and so. . . . But a schema is not assessable as true or false, nor is any state whose "content" can be given only in schematic terms assessable as true or false (p. 202).

One can imagine a version of the dual component theory of which it might be said that it differs only terminologically from Evans's position. Suppose we reserve "thought" for something which essentially has truth conditions. The dual-component theorist of the kind I envisage will agree with Evans that particular-thoughts are Russellian. Suppose we reserve "thought\*" for the intrinsically mental component dear to the dual component theorist. Let the theorist agree that thoughts\* are not representations (in Evans's sense), but only schemata; that they do not themselves essentially possess truth

conditions, though in favourable circumstances they may be associated with truth conditions. Thoughts\* are evidently non-Russellian. Can Evans agree with the dual component theorist that there exist thoughts\*? Indeed, that anyone who has a thought must have a thought\*?

Consider a case in which a subject, *S*, thinks, on earth, of an object, *a*, that it is *F*. His Doppelgänger on Twin Earth, *S'*, thinks of *a*'s Doppelgänger, *a'*, that it is *F*. Evans can, and does, allow not only that *S* and *S'* are in a similar condition (there is a *general* disposition which each possesses), but also that the respect in which they differ (their thinking different thoughts) "results from" (p. 204) this similar condition: the general disposition, together with the distinct environment, yields the different specific conditions, one *a*-related, the other *a'*-related. So it would seem that there should be no quarrel between Evans and this brand of dual component theorist. (One should also bear in mind that it is no part of Evans's position that the mind of one who essays, but, through absence of any appropriate object, fails to attain, a particular-thought is wholly vacant (p. 45). Indeed, Evans allows (e.g. p. 203) that for such a person it will be just as if he were having a particular-thought.)

Substantial disagreement would occur if the dual component theorist were to hold that only thoughts\* and not thoughts count as genuinely *mental*. Against the view that all genuinely mental states are intrinsic to the subject, Evans has an argument that I think is both illuminating and decisive (p. 203). Let it be agreed that an adequate explanation of actions can be given entirely in terms of the agent's mental states. Actions, however, the explicanda, are often singular. We want to know why a subject greeted *that person*, and not merely why he greeted someone having certain general characteristics; why he blamed *himself*, and not merely why he blamed a person having certain general characteristics; and so on. We must explain not merely why *S* reacted so violently to an object having the general features shared by *a* and *a'*, but why he reacted to *a* as opposed to *a'*. (Answer: he could see *a* but could not see *a'*.) The need for Russellian thoughts in our psychological descriptions is correlative with our need to explain *object-related behaviour* in the way that we do. Perhaps this is not the pattern of explanation sought in natural science. If there is a biological explanation of human behaviour, perhaps it will involve only intrinsic states of organisms. If so, this is because it will describe the explicanda in general terms. This would be simply one more reason for seeing biological explanations as quite distinct in kind and purpose from our everyday explanations in terms of mental states. It is no reason for holding that all genuinely *mental* states are intrinsic.

### 3 Existence

In fiction, and also in less systematic and self-conscious cases of pretence, expressions which appear to be singular terms are used in what appears to be a non-Russellian way. "Hamlet" and "that little green man" appear to be

singular terms, and yet also appear to be usable in make-believe contexts to express thoughts even though these involve no (real) Hamlet, and no (real) little green man. An opponent of Evans's account may take both appearances at face value, saying that the expressions are, even in these contexts, being genuinely used as singular terms, but non-Russellian ones.

The problem comes into especial prominence in the case of true singular negative existential sentences. Taking the sentences as genuinely singular (as opposed to being denials of existential generalizations) seems to force us to regard the apparent singular terms as being used really as singular terms. Taking the sentences as true seems to involve regarding such apparent singular terms as non-Russellian.

The general pattern of Evans's response is as follows. The fictional use of apparently non-Russellian singular terms is not inconsistent with the overall Russellian conception. For the fictional use is typically not a real use, in which a genuine thought is expressed, and the possibility of genuine understanding created, but only a pretend use, in which it is pretended that a thought is expressed, and the possibility created is not of genuine understanding, but only of an imitation of understanding which Evans calls "quasi-understanding". In their fictional use, their role in the pretence is the fictional image of the role of Russellian singular terms used in non-fictional contexts: whatever properties an expression of that category (say a proper name or a complex demonstrative) would have in non-fictional contexts are mimicked in fictional ones, including such properties as being Russellian, being information-linked, and so on.

To have a counterexample to the claim of the Russellian status of the kind of singular terms with which Evans is concerned, we would need to find an expression of the kind that was a singular term but not Russellian. If we view expressions like "Hamlet", which are supposed to be counterexamples, from outside the fiction, then we must agree that they are non-Russellian but not that they are singular terms. If we view them from within the fiction, we must agree that they are (make-believable) singular terms; but they are also (make-believable) Russellian.

This does not address the problem of negative existentials, for here the (apparent) singular terms are not being used in a fictional context: the problematic sentences are problematic precisely because they are genuinely true.

Various features of audience reaction to fiction have to be understood as involving a *serious* exploitation of it, serious in that what is said or thought can be properly assessed for genuine (and not just fictional) truth or falsity. Evans's treatment of singular negative existentials sees them as involving such a serious use.

The general idea is that someone who utters such a sentence should be likened to someone who makes a move within a pretence in order to express the fact that it is a pretence. He is not like someone who tries to

prevent a theatre audience from being too carried away by jumping up on the stage and saying: “Look, these men are only actors, and there is no scaffold or buildings here – there are only props.” Rather, he is like someone who jumps up on the stage and says: “Look, Suzanne and the thief over there are only characters in a play, and this scaffold and these buildings are just props.” The audience must be engaged, or be prepared to engage, in the make-believe, in order to understand what he is saying (p. 369).

In working this out in a little more detail, let us adopt Evans’s abbreviation of “it is make-believedly the case that ( )” as “\*( )\*”. A sentence with a use within a pretence has the following property: \*it has a truth-condition (expresses a proposition)\*. But, Evans implies, it typically will not have a truth condition: typically, that is, it will not really express a proposition.<sup>10</sup> (As we shall see, there is a vital exception to what is typical.)

Evans sees negative existentials as implicitly involving the expression “really”, of which he gives the following account:

“Really” is a word which, when prefixed to a sentence, produces a sentence such that an utterance of it is true (absolutely) if and only if the sentence preceded by “really” is itself such that there is a proposition expressed by it when it is uttered as a move in the relevant game of make-believe, and this proposition is true (absolutely) – not merely \*true\* (p. 370).

A negative existential sentence, say “Hamlet does not exist”, is to be seen as having the following logical form: it is not the case that really (Hamlet exists). Forgetting for a moment the atypical case, the standard case is one in which, as in the present example, the contained positive existential does *not* express a proposition when it is uttered in the “relevant game of make-believe”. That is, it is not the case that “Hamlet exists” has a genuine and non-fictional truth-condition. (We will shortly look at a justification for this position.) Hence, by the truth-condition for “really”, “really Hamlet exists” is not true, and so its negation is true, which is the desired result.

The atypical case is that in which an utterance, though used fictionally, in fact really expresses a proposition. This is the phenomenon which Evans labels the “game-to-reality shift” (Ch. 10.2 and pp. 370–1). Suppose that *X* is seeing a little green man, but believes that he is hallucinating. His utterances of the form “That little green man is *F*”, though intended as part of a make-believe, in fact refer to the little green man he is seeing: the game shifts to reality. In particular, “that little green man exists” does express a proposition,

10 For understanding Evans on this point I am particularly indebted to Martin Davies.

indeed a true proposition. So “Really, that little green man exists” is true, and its negation false, which is again the desired result.

If we were considering only existential sentences, the truth condition for “really” could be simpler: we get a truth if the sentence to which it is attached has a truth-condition as well as a \*truth-condition\* and a falsehood if, while having a \*truth-condition\* it lacks a truth condition. But “really” can attach to other forms of sentence, e.g. “That little green man is really bald”. In the case envisaged in the last paragraph, in which we have the game-to-reality shift, the truth of such an utterance in such a context will require that the sentence “that little green man is bald” express a proposition, but this by itself is not enough: in addition, the proposition expressed must be true.

Evans contrasts his account of “really” with another simpler one, which would be available if all make-believe were existentially conservative, that is, involved only pretence concerning actual objects. Then “really” could simply indicate that what followed was to be assessed for genuine truth or falsehood, as opposed to fictional truth or falsehood. But this will not do as things stand. For when a fiction \*introduces an entity\*, say Hamlet, a sentence which follows “really” and which \*mentions this object\* is not assessable for genuine truth or falsehood: such a sentence has no truth condition but only a \*truth condition\*; it does not really express a proposition. (Here there is no game-to-reality shift.) So the simpler account would, incorrectly, entail that the whole sentence, say, “really, Hamlet exists”, says nothing.

Evans does not himself offer any explicit justification for the view that fictional sentences containing apparent singular terms typically express no proposition. On the face of it, it might seem that, for any fiction, we can represent to ourselves what it would be for it to be really true. So it might seem that we are committed to the view that fictional sentences do genuinely express propositions, ones which we can envisage being true (absolutely).

Against this position, Evans could avail himself of a line of argument used by Kripke (1973), in discussing whether unicorns could have existed. We have to ask: what makes the envisaged counterfactual situation one in which *unicorns* exist, as opposed to one in which there are creatures very like unicorns, but not really unicorns? Kripke says that this question cannot be answered, and that the moral is that no possible situation counts as one in which unicorns exist.

Let us try applying this to a particular fictional name, say “Hamlet”. The anti-Evans claim is that, though we know that this is merely a fiction, we know what it would be for the sentences in the story to be true. So the sentences must express genuine propositions, evaluable for genuine truth. What one cannot deny is that this holds for the generalizations in the story: once upon a time there was a prince who. . . . But we can deny that it holds also for the singular sentences, e.g. “Hamlet was indecisive”. For what would it be for

this to be really true? It is not enough that there be an indecisive person whose father was a king and was murdered, and who. . . . For, as we are inclined to say, a person might have done all those things without being Hamlet. But then what condition is sufficient to discriminate between situations containing someone rather like Hamlet and situations containing Hamlet himself? It seems that no condition is sufficient. If this is accepted, the moral is that we cannot, after all, represent to ourselves what it would have been for “Hamlet was indecisive” to be genuinely, and not just fictionally, true. So sentences of the envisaged kind have a \*truth-condition\* but no truth-condition. They do not express genuine propositions.

Evans says that one of the conditions for an adequate account of negative existentials is that they must not be represented as metalinguistic (pp. 344–5). As he says, it is obvious that a singular term appearing in such a sentence is used and not merely mentioned. Understanding what a negative existential says requires familiarity with the relevant use of the singular term. But a metalinguistic account, which, in its crudest form, analyses “*a* does not exist” as “‘*a*’ does not refer”, does not meet this condition. Connected with the point that the singular term is used, and not merely mentioned, is the point that the significance of the whole sentence must depend upon the significance of the parts used to compose it. Rival theories, for example Kripke’s, are criticized for their failure to give any explanation of how the positive existential component of negative existentials can be genuinely significant, and thus contribute to the significance of the whole (cf. pp. 349–51). It is instructive to see how Evans’s theory seeks to meet these conditions.

The proof that on Evans’s account the apparent singular term is used and not merely mentioned in a negative existential is this: understanding a sentence of the form “Really S” involves quasi-understanding S, and this in turn involves familiarity with the game of make-believe within which S has its \*truth condition\*. (Seeing the logical form of negative existential truths as involving “really” is crucial to this point.) No such requirement could hold if the apparent singular term in S were merely mentioned. Moreover the sense of a negative existential is sensitive to the sense, or rather, perhaps, quasi-sense, of the apparent singular term. Indeed, two \*co-referring singular terms\* could, on this account, \*express different thoughts\*.

Has the requirement that the sense of a compound is a function of the sense of its parts been met in a properly full-blooded way? Is not the notion of quasi-sense, which I have introduced as the correlate of Evans’s quasi-understanding, merely papering over the fact that the positive existential component of a true negative existential is not contributing in the proper way to the sense of the whole? Is it legitimate to allow quasi-sense to contribute to the determination of sense? That the answers to these questions fall out in favour of Evans’s account is, I think, established by the mere fact (assuming it to be a fact) that sentences like “I admire Hamlet” are coherent and wear their logical form on their sleeve. For here the fictional expression, possessing

(on Evans's account) only quasi-sense, contributes to genuine sense. This is what Evans has in mind by the *serious* use of fictional expressions, and it must be stressed that his account of negative existentials is preceded by a *general* discussion of serious use.

In a sentence like "Vulcan does not exist" what is the make-believe in virtue of which "Vulcan exists" has a \*truth condition\*? Evans does not say. There is a choice. We might consider the benighted astronomers as having gone in for unwitting make-believe. Their illusion of understanding, in using "Vulcan", was systematic enough to make this just like a fiction, except for their lack of suitable intentions and of conscious awareness of what was going on. Alternatively, we might think that it is we, who know that Vulcan does not exist, who create the make-believe, perhaps in the very utterance in which we claim it to be mere make-believe. Evans's account stands in need of completion at this point.

One aspect of the account that is likely to attract criticism is the idea of the game-to-reality shift. Indeed, a critic might use an *ad hominem* argument: the possibility of the shift, it might be alleged, is inconsistent with Evans's insistence that a speaker can only refer to what he intends to refer to (cf. p. 318). I suspect that these criticisms can be deflected. But let me stress the dependence of Evans's account on the game-to-reality shift. Were the shift disallowed, he would have no explanation of the truth of "that little green man really exists", as uttered in the circumstances in which a subject sees a little green man but takes himself to be hallucinating. If there is no shift, the sentence comes out false, by the truth condition for "really".

I am optimistic that Evans's account of existence can be completed into a correct account. It is the only account of existence that has any plausibility at all when it comes to explaining the possibility of truths of the form "that does not (really) exist". (The possibility is realized within the context of a hallucination.) This is worth stressing, since many who, like Russell, would think that "Vulcan" can be treated as a description, permitting a quantificational treatment of "Vulcan does not exist", would not contemplate such a course for demonstratives.

#### 4 Coda

I have discussed only the main plot of *The Varieties of Reference*. I have not done justice to how much this remarkable book contains. It is packed with intense and sustained argument and is brim full of ideas on a wide variety of subjects. These ideas are sometimes casually thrown out (like the footnote (p. 104, n. 22) suggesting a principled way of distinguishing thinkers from other information-processors), and sometimes belong in detailed discussions which deserve attention independently of their contribution to the main theme (for example, the chapter on the self, and that on recognition). The ore lies deep, but it is extremely rich, and amply repays the effort needed to extract it.



There is no doubt that the book will be heavily mined, and will be, deservedly, central to discussions of reference and singular thought for years to come. It not only justifies John McDowell's editorial hope that it will enable us to appreciate how much philosophy has lost by Evans's early death: it also shows us how much philosophy has gained by his life.<sup>11</sup>

11 Evans's book was the subject of a series of seminars held at Birkbeck College in the Summer Term, 1983. I am grateful to those who participated for any ideas I may have borrowed. I would like to thank the following for helpful comments on an earlier draft: Martin Davies, Marcus Giaquinto, Samuel Guttenplan, Gary Jenkins, Hans Kamp, Ian McFetridge, Colin McGinn, Anthony O'Hear, Anthony Savile.

# Concepts without boundaries

---

## I Classification by classes

Some concepts classify by setting boundaries but some do not. In the philosophical tradition, the former have received all the attention, and have lent a distinctive character to attempts to study classificatory concepts and their linguistic correlates. Within what I shall call the “classical picture”, a picture which dominates most thinking about thought and language, there is no room for the thesis I wish to put forward: that concepts can classify without setting boundaries.

According to this classical picture, the job of classificatory concepts is to sort or segregate things into *classes* by providing a system of pigeon-holes, by placing a grid over reality, by demarcating areas of logical space. Boundaries are what count, for a concept must use a boundary to segregate the things which fall under it from the things which do not. This intuitive view receives definitive expression in classical semantic theories. A predicate, linguistic vehicle of a concept, is thought of as having a meaning which fixes its extension, the set of things of which it is true. A semantic theory will provide an at least partial characterization of a predicate’s meaning by specifying this set in some appropriately revealing fashion. In the light of such specifications, one can model the logical features of the language in which the predicate occurs by generalizing over the sets which predicates do, or can, determine.

Thus the classical picture, informed by a connection between concepts and sets present in the very word “classify”, sees the theoretical resources of set theory as the proper instruments for describing language and thought.

Classes, and sets, have sharp boundaries. Hence, at least *prima facie*, a description of concepts or predicates in terms of what sets they determine is a description of them as boundary-drawers. This mode of description, and the picture which underlies it, thus makes no room for concepts without boundaries, those which are not boundary-drawers. We have a choice: we could take the classical picture as exhausting the ways classificatory concepts can be, and conclude that there are no concepts without boundaries; or else, convinced that there are such concepts, we could reject the comprehensiveness of the classical

picture. What I suggest is that almost all concepts lack boundaries, so that the classical picture is of very little use to us.

The concepts which classify without setting boundaries include the ones traditionally counted as vague: *red*, *heap*, *child*, *bald*, to take some famous examples. The first stage in showing that these are boundaryless concepts involves showing that there is no set of which they are true: they do not classify at all, if the only way to classify is to assign things to classes or sets.

## 2 There is no set of red things

Sets have sharp boundaries, or, if you prefer, are sharp objects: for any set, and any object, either the object quite definitely belongs to the set or else it quite definitely does not. Suppose there were a set of things of which “red” is true: it would be the set of red things. However, “red” is vague: there are objects of which it is neither the case that “red” is (definitely) true nor the case that “red” is (definitely) not true. Such an object would neither definitely belong to the set of red things nor definitely fail to belong to this set. But this is impossible, by the very nature of sets. Hence there is no set of red things.

This seems to me as certain as anything in philosophy, yet it can often be a bitter pill to swallow, by non-philosophers and by philosophers alike. In some debates about abortion, one can feel a real sense of shock at the realization that there is no set of persons: the concept *person* is vague at just the relevant point. The difficulty is that moral concepts are often boundary-drawing (especially so the more naive the morality), and legal concepts typically have to be. Trying to tie the application of a boundary-drawing concept (as *may legitimately be aborted* is supposed to be) with a boundaryless one like *is a person* poses a problem which is simply not soluble in the straightforward terms in which it is often posed. A quite general reflection on boundaries, their absence and presence, should reshape what one could expect to emerge from a discussion of abortion and the law.

Philosophers, too, are attracted by the classical picture, even when not engaged upon formal semantic projects. To take one example from a million, when Peter Strawson some time ago considered whether “Socrates is human” is equivalent to “Socrates belongs to the set of humans” his many interesting observations did not include the point that the equivalence must fail since there is no set of humans.

If vague predicates and vague concepts do not have “extensions” – sets of things of which they are true – they do not draw boundaries, at least not in any simple sense. For a boundary should divide things into two sets, those which fall on one side and those which fall on the other. So if vague predicates do not effect a division into sets, they draw no boundaries.

But, it may be objected, this is too simple. For may not a vague predicate draw an unsharp boundary? May not an unsharp boundary work without

dividing into two? May there not be things whose status is unresolved by the boundary?

When one says that a vague predicate does not draw *sharp* boundaries, “sharp”, I believe, does no work, for there is only one kind of boundary. Hence we cannot regard vague predicates as drawers of boundaries, but ones which are unsharp. I shall establish this by the following route. Anything worthy of the name *boundary* will effect set-theoretically describable divisions, even if more complex ones than the simple twofold division envisaged just now. But any such division, however complex, will misdescribe the functioning of a vague predicate.

### 3 An unsharp boundary is many sharp boundaries

I shall start with a general reason for thinking that there is no adequate set-theoretic description of vague predicates. It is that a set-theoretic description of a language typically ends up identifying a set of truths. (More exactly, it specifies for each sentence a condition upon which the sentence will belong to this set, or to the set of truths-upon-*S*, for a relativization to some structure, *S*.) But the argument which showed that there was no set of red things shows also that there is no set of truths in a language which contains the predicate “red”. For consider a sentence ascribing *red* to a borderline case. This sentence will neither (definitely) belong nor (definitely) fail to belong to the set of such truths; which is another way of saying that there is no such set.

But surely (one may object) this simple thought does not do justice to all the cunning twists and turns which set-theorists have made in their attempts to describe vagueness. What about fuzzy logic? What about supervaluation theory?

I believe that the simple minded point of a moment back – that there is no set of truths – does ultimately carry the day against such theories. But I recognize that to convince an opponent involves more detail.

The preliminary claim I need to make is that one cannot do justice to the phenomena of vagueness, in particular to phenomena called “higher order vagueness”, simply by increasing the number of sets of individuals associated with a predicate. A set-theoretic description might start by associating a vague predicate not with two sets but with three: the set of things of which the predicate is (definitely) true, the set of things of which it is (definitely) false, and the remainder, the set of borderline cases. So far, so good: a sharp predicate has two extensions, a positive extension, and a negative one (its complement within the domain) whereas a vague predicate has three, a positive one, a negative one, and a penumbral one (the complement within the domain of the union of the other two).

But a predicate which effects such a threefold partition is not vague. This fact, which shows why one cannot characterize vagueness merely in terms of borderline cases, follows from the fact that the partition does identify a set of

truths, which we have seen to be inconsistent with vagueness. Once again more detail will reinforce the point. Consider a child developing into adulthood. We cannot associate “child” with a twofold division because there is no set of children (at a time). There is no such set because of the borderline cases. Thus far, some encouragement for a threefold partition, which explicitly allows for borderlines. However, essentially the very point which scuppers a twofold partition also scuppers a threefold one: the latter posits a set of children, or at least a set of definite children. Yet it would be as absurd to suppose that a heartbeat could make the difference between membership of this set, and consignment to the set of borderline cases, as it would be to suppose that a heartbeat could make the difference between belonging to the set of children and belonging to the set of non-children. Childhood, even definite childhood, fades gradually away, and does not come to a sudden end. This is not to deny that we can by convention stipulate a sharp boundary; it is only to say that our concept *child* does not supply one.

A proponent of set-theoretic divisions might seek to meet this point by making more divisions. He might say that a predicate is sharp, that is, not vague at all, or as I shall say is vague<sub>0</sub>, just on condition that it draws a single boundary, thus partitioning the domain into two sets; that a predicate possesses the lowest level of vagueness, is vague<sub>1</sub>, just on condition that it draws two boundaries, partitioning the domain into three sets. If it is unacceptable to suppose that “child” or “red” does this (on account of the unacceptability of supposing that there is a last heartbeat of one’s childhood, or of one’s definite childhood), then the correct description of such a predicate must look further up the hierarchy: perhaps it draws four or forty boundaries, making correspondingly many partitions of the domain.

The generalization of this set-theoretic approach is that a predicate is vague, iff it draws  $2^n$  boundaries, thus partitioning the domain into  $2^n + 1$  sets.\* A predicate is sharp iff it is vague<sub>0</sub>; is vague iff it is vague<sub>n</sub> for some positive  $n$ ; is higher order vague iff it is vague<sub>n</sub> for some  $n > 1$ , and is radically vague iff it is vague<sub>n</sub> for no  $n$ . The phenomena mentioned a moment back – such facts, now agreed by the set-theorist, as that there is no sharp division between children and borderline cases of children – are to be described, on the envisaged approach, by going higher in the hierarchy of higher order vagueness. If the set-theoretic description assigns a cut-off where none can be found in the actual use of the predicates, then the description has not set the level of vagueness high enough. The hope is that the unlimited upwards mobility is enough to enable one to get on top of all the phenomena.

This hope, however, is groundless. Indeed, its very structure should be unappealing: you do not improve a bad idea by iterating it. In more detail,

\* [Added note] Williamson [1999] has shown that there are other ways of using set-theory in describing higher-order vagueness.

suppose we have a finished account of a predicate, associating it with some possibly infinite number of boundaries, and some possibly infinite number of sets. Given the aims of the description, we must be able to organize the sets in the following threefold way: one of them is the set supposedly corresponding to the things of which the predicate is absolutely definitely and unimpugnably true, the things to which the predicate's application is untainted by the shadow of vagueness; one of them is the set supposedly corresponding to the things of which the predicate is absolutely definitely and unimpugnably false, the things to which the predicate's non-application is untainted by the shadow of vagueness; the union of the remaining sets would supposedly correspond to one or another kind of borderline case. So the old problem re-emerges: no sharp cut-off to the shadow of vagueness is marked in our linguistic practice, so to attribute it to the predicate is to misdescribe it.

#### 4 Fuzzy logic and supervaluations

In effect, this same point is what scuppers the set-theoretic descriptions of vague languages offered by fuzzy logicians and by supervaluations theorists.

The fuzzy logic I envisage associates with each predicate not a set whose only members are individuals, the individuals of which the predicate is true, but what Goguen (1969) calls a "J-set": a function from each object in the domain of discourse to a real number in the closed interval  $[0,1]$ . The number, as value to the function, represents the degree to which the predicate is true of the object which is argument to the function. The real numbers are continuous, so surely now we have a mode of description apt for vagueness, one which does not necessitate sharp boundaries? Yet a fuzzy set is a genuine set, a completely sharp object.

The reason for thinking this approach cannot succeed is essentially the same as that given earlier: the fuzzy logician, too, will (whether he likes it or not) be committed to a threefold partition: the sentences which are true to degree 1, those true to degree 0, and the remainder. But to what in our actual use of language does this division correspond? It looks as if, as before, it should correspond to the sentences true beyond the shadow of vagueness, those in some kind of borderline position, and those false beyond the shadow. But, as several times noted, we do not know, cannot know, and do not need to know these supposed boundaries to use language correctly. Hence they cannot be included in a correct description of our language.

Fuzzy logic does, indeed, describe a feature of our use of many vague predicates: that, and how, they are associated with a dimension of comparison. Our use of "red" is properly regarded as regulated by such principles as that anything redder than a red thing is red. The "redder than" relation is tracked by fuzzy logic's numerical ordering, which in turn bears straightforwardly on the applicability of "red". For many predicates, things are more complex: many relations are relevant to applicability. Childhood is affected by

age, but also by many other factors as well, so that two individuals of the same age may differ in point of how much of a child they are. Yet, plainly, fuzzy logic could be supplied with the resources to describe the nature and weights of a whole complex of applicability-determining relations. Could we not somehow reap these benefits without succumbing to fuzzy logic's three-fold partitioning?

Here is one possible line of thought. One can expect the empirical data for fuzzy logic to be quite messy. One might run trials in which one asked people to do two kinds of thing: order all the objects in terms of their strength as candidates for application of the predicate; and identify the definite cases and the definite non-cases. The results would be variable, both intra- and inter-personally. Let us say that an *admissible* J-set for a predicate is one which matches some trial for that predicate both in point of order and in point of definite cases. That is to say, if in the trial the subject treats  $\alpha$  as a better case of the predicate than  $\beta$ , then the J-value for  $\alpha$  must exceed that for  $\beta$ ; if the subject identifies an object as a definite case for the predicate, the J-value for that object must be 1; if the subject identifies an object as a definite non-case for the predicate, the J-value for that object must be 0. When it comes to specifying truth, the theorist could adopt a vague definition, for example he could say that an atom  $\phi\alpha$  is true iff the object denoted by  $\alpha$  has J-value 1 for almost all J-sets admissible for  $\phi$ .

The general idea behind the strategy is to take the set-theoretic description as a kind of basis, and exploit it in a vague way to deliver an account of whatever one takes to be the central semantic notion.

The idea is not unattractive, but it does not fall within the scope of the approach I wish to attack. For what is envisaged is that the real work of describing the functioning of a predicate is done not by fuzzy logic itself, but in terms of some *vaguely* specified semantic notion. The proposal, then, ends up as not one in which a predicate is described by being associated with an extension or with boundaries.

A similar series of moves can be made in connection with supervaluations. The supervaluational theory, too, will end up making a threefold partition: the set of sentences true-upon-all-sharpenings, the set of sentences false-upon-all-sharpenings, and the remainder. An attempt to do justice to "higher order" vagueness by acknowledging vagueness in the notion of a sharpening would force one outside the set theoretic language in which the theory is supposed to be couched, and would mean that the real work of semantic description was being done in a vague language, rather than in a set-assigning one.

## 5 Boundaryless concepts

A vague concept is boundaryless in that no boundary marks the things which fall under it from the things which do not, and no boundary marks the things

which definitely fall under it from those which do not definitely do so; and so on. Manifestations are the unwillingness of knowing subjects to draw any such boundaries, the cognitive impossibility of identifying such boundaries, and the needlessness and even disutility of such boundaries.

To characterize a vague concept as boundaryless is an improvement on characterizing it as one which permits borderline cases, since a non-vague concept may admit borderline cases. If “child\*” is defined as true of just those people whose hearts have beat less than a million times, false of those whose hearts have beaten more than a million and fifty times, and borderline with respect to the remaining people, it has borderline cases but behaves quite unlike our paradigms of vagueness.

A boundaryless concept cannot be described in set-theoretic terms. How can it have a classificatory role? How is it to be described, either semantically or in terms of cognitive processing? How are the paradoxes of vagueness to be avoided?

Scepticism about whether boundaryless classification is possible can be set to rest, I believe, by contemplating a very familiar case: the colour spectrum, as displayed, for example, in an illustration in a book on colour. Looking carefully, we can discern no boundaries between the different colours: they stand out as clearly different, yet there are no sharp divisions. There are bands, but no bounds. This does nothing to impede the classificatory process: the spectrum is a paradigm of classification.

The image of pigeon-holes is powerful. Is there a comparable one which would represent how boundaryless concepts classify? We could, perhaps, think of such concepts as like magnetic poles exerting various degrees of influence: some objects cluster firmly to one pole, some to another, and some, though sensitive to the forces, join no cluster.

At least one aspect of this image deserves more literal statement. Boundaryless concepts tend to come in systems of contraries: opposed pairs like child /adult, hot/cold, weak/strong, true/false, and the more complex systems exemplified by our colour terms. This is a natural upshot of boundarylessness, as we can see by reflecting on what is involved in grasping a concept.

Such a grasp, it must be agreed on all sides, involves knowing how something would have to be for the concept to apply to it, and how something would have to be for the concept not to apply. A distinctive feature of the classical picture is that it sees here a single ability. Grasping what a concept excludes is part of grasping the concept, and is achieved through the mediation of no other non-logical concept. Hence it is very natural to see the division between what a concept includes and what it excludes in terms of a boundary. Certainly, perception of a boundary would be enough; but the proponent of boundarylessness will insist that it is not the only way.

On the alternative picture, what a concept excludes is graspable in a positive way, mediated by other contrary concepts. A grasp of *red* attains grasp of



what is not red at a derivative level, via a grasp of *yellow*, *green*, *blue* and so on. A system of such concepts is grasped as a whole, as can be seen in the way paradigms are used in learning. There are paradigms of red, but nothing is non-derivatively classifiable as a paradigm of not-red. Any paradigm of another colour will serve as a paradigm of how not to be red, but only in virtue of its positive classification as another colour.

Not just any clear case of the non-applicability of a concept will serve to help a learner see what the concept excludes. Television sets, mountains and French horns are all absolutely definite cases of non-children; but only the contrast with *adult* will help the learner grasp what *child* excludes. So it is no accident that boundaryless concepts come in groups of contraries. Correlatively, the image of attracting poles, replacing the classical image of pigeon-holes, is not without value.

It also serves to record some empirical data. For example, subjects asked to classify a range of test objects using just “young” and “old” make different assignments to these words from those they make to them when asked to classify using, in addition, “middle-aged”. The introduction of a third magnetic pole can attract some of the things only loosely attached to two existing ones, without diminishing the forces the existing ones exert.

## 6 Paradox

Let me now turn to the question about paradox, which might take a more aggressive form: does not the very notion of boundarylessness make the paradoxes unavoidable? For the absence of a boundary has been treated as the impossibility of very similar things differing in point of the applicability of a predicate. But then it seems that we can form a sorites series of objects, adjacent pairs being too similar to merit a difference in applicability, but remote pairs being sufficiently dissimilar to require a difference. Starting with a clear case of red, we assemble closely resembling patches in a series through which the colour shifts gradually towards yellow. The boundarylessness of red is supposed to ensure that there is no adjacent pair of which the first is red and the second not. The first member, by hypothesis, is red. Hence, by boundarylessness, the one adjacent to it is also red; and so on. So does not familiar reasoning lead inexorably to the intolerable conclusion that all members of the series are red, even the yellow ones?

This worry can take a form which can only be assuaged by a technical and formal semantic theory, and nothing of that kind is on offer here. However, let me make a simple observation which should establish that the present picture of vagueness as boundarylessness is no less well placed than any other to come to terms with sorites reasoning.

The classical picture has a totalitarian aspect: there is no difference between its being not mandatory to apply a concept and its being mandatory not to apply it. If the very nature of the concept *prime*, together with the

nature of some number, say eight, does not require you to apply the concept to it, then the very nature of the concept, together with the nature of the number, requires you not to apply the concept to it. For a rational and fully informed thinker, there is no freedom.

By contrast, vagueness offers freedom. It can be permissible to draw a line even where it is not mandatory to do so. No one can criticize an art materials shop for organizing its tubes of paints on various shelves, including one labelled “red” and another “yellow”, even though there is a barely detectable, or perhaps even in normal circumstances undetectable, difference between the reddest paint on the shelf marked “yellow” and the yellowest paint on the shelf marked “red”. Hence one can consistently combine the following: *red* draws no boundaries, that is, there is no adjacent pair in the series of tubes of paint such that the nature of the concept, together with the colour in the tube, requires one to apply *red* to one member of the pair but withhold it from the other; yet one can draw a boundary to the reds, that is, one may behave consistently with the nature of the concept in drawing a line between adjacent pairs.

The envisaged attack on boundarylessness can be set out as the following argument, which makes plain how the recent observation addresses it. A boundaryless concept is one which, for closely similar pairs, never makes it mandatory to apply the concept to one member of the pair, and withhold it from the other; hence, the argument runs, a boundaryless concept is one which, for closely similar pairs, makes it mandatory never to apply the concept to one member of the pair, and withhold it from the other. The inference depends upon the move from something being not mandatory to its being forbidden; a move legitimate within the totalitarianism of boundary-drawing concepts, but not within the liberality of boundarylessness.

I do not suggest that this simple observation puts an end to the lure of sorites reasoning, which, like a virus, will tend to evolve a resistant strain. Must there not be an outer limit to the things to which it is mandatory to apply “red”, and a first member of the sorites series with respect to which we have licence to withhold? The answer is “No: ‘mandatory’, too, is boundaryless”; though I shall not now stop to show how this answer can be justified. It is enough for the moment to have shown that boundarylessness should give no special encouragement to paradoxical sorites-style reasoning.

## 7 Semantics for boundarylessness

If standard set-theoretic descriptions are incorrect for boundaryless concepts, what kind of semantics are appropriate? A generalization of the considerations so far suggests that there is no precise description of vagueness. So what kind of description should be offered? More pointedly, I hear a certain kind of objector say: we can’t even tell what boundarylessness is until you give us your semantics.

If driven in this way, I would urge an idea of Donald Davidson's. A semantic theory can quite legitimately be *homophonic*, that is, can reuse in the metalanguage the very expressions whose object-language behaviour it is attempting to characterize. Asked how a boundaryless predicate like "red" works, my first response would be: "red" is true of something iff that thing is red.

Whether or not vagueness is at stake, this Davidsonian idea has met with resistance. No one could claim that such remarks are untrue, but they have been held to be trivial or unilluminating. Many such objections are confused, or are based on a misunderstanding of what, by Davidson's lights, a semantic theory should aim to achieve. In his view it should enable us to understand how some expression conspires with others to fix a truth condition, an understanding which would answer the question: "what are these familiar words doing here?"; and it should supply an account of what it would be enough for a speaker to know, in order to understand a language. Homophony impedes neither aim. On the contrary, it can provide a check upon their successful accomplishment. But a homophonic semantics for vague expressions could lead to two more specific objections, one misguided, but one sound and important.

The misguided one is that homophonic semantics will fail to make explicit which predicates are vague. In fuzzy logic, for example, the distinction will emerge in the structure of J-sets. Those associated with vague predicates will, for some or many objects as arguments, deliver numbers intermediate between 0 and 1 as values; whereas a sharp predicate's J-set will have 0 or 1 as the only values, whatever objects are arguments. In a homophonic theory the information is in a way present, but is inexplicit. It is present, because in specifying the applicability of, say, "bald" homophonically, you specify it vaguely and so *as* vague. But it is inexplicit, since there is no rule, usable by one who did not yet know which predicates are vague, on the basis of which, together with the semantic theory, he could pick out the vague ones.

However, I know of no reason why this fact in itself counts against the homophonic approach. First, I know no reason for thinking that all such information must be made explicit, if the mentioned aims of semantics are to be achieved. Secondly, even if the mentioned aims, or others, did require that the information be made explicit, I know of no argument to establish that this cannot be achieved simply by means of a list, and thus consistently with homophony.

The sound and important objection is that the homophonic approach fails us in connection with logic. Davidson himself envisages a first-order metalanguage, and thus a metalanguage of which classical model theory is true, and thus a metalanguage in which predicates are associated with sets as their extensions. Thus envisaged, the project succeeds in fixing a logic for the object language, namely, classical first order logic; but it fixes it while at the same

time mischaracterizing the semantics of the object language. For, despite syntactic homophony, if the metalanguage is first order, its predicates will be boundary-drawing, and so will misrepresent the object language predicates as also being boundary-drawing.

This first order feature of Davidson's proposals is, as he says, inessential. Abandoning it makes possible serious homophony: an account of the object language predicates in which they are not merely reused in point of their sounds or shapes, but also in point of their meaning. The problem then, however, is to say something worthwhile about the logic of the object language. There are two obstacles. First, we do not know what our actual logic, which would be reapplied homophonically, is. We do not know, for example, whether every instance of *P* or *not-P* is counted true in our language and thought, and one pertinent reason for this doubt stems from vagueness. Secondly, even if we knew what our actual logic is, we could not uncritically reuse it in a semantic project, for the existence of sorites reasoning casts doubt upon whether we are right to subscribe to the logic to which we actually subscribe.

The logic of vagueness, characterized as boundarylessness, thus remains to be described. I believe that the way forward involves taking the notion of a vague object as basic; but this is a suggestion I shall not pursue here.

## 8 Paradigms and prototypes

If the semantic description of a vague concept is to have all the thinness of homophony, can we not achieve a richer description in other ways, perhaps in terms of the psychological mechanisms whereby a vague concept is acquired or applied?

For example, it seems that very often a boundaryless concept is acquired on the basis of paradigms. We acquire the concept from the inside, working outwards from central cases, and locating the central cases of contrary concepts, rather than from the outside, identifying boundaries and moving inwards. Can this thought be used to say anything illuminating about the nature of boundaryless concepts?

If it can, I am not sure how. Perhaps we should try to specify a boundaryless concept's relation to the world in terms of a paradigm – an object,  $\alpha$ , to which it quite definitely applies, and which might therefore be an appropriate example to use in a teaching situation – together with a relation of similarity. Then we might say, non-homophonically, that the concept is true of something iff that thing is sufficiently similar to  $\alpha$ .

But the suggestion has many vices. For one, it presupposes, without any justification, that every boundaryless concept must be instantiated. For another, the condition will not state anything which users of the concept have to know, since there may be more than one paradigm, and so one could master the concept just as well without knowing anything of  $\alpha$ . Thirdly,

anything which might have worried one about boundarylessness, for example any problems about the sorites paradox, naturally remain just as they were, though now attaching to the similarity relation.

Surely, however, I should take note of, and perhaps make use of, psychologists who have, it might be thought, investigated essentially this phenomenon under the name of prototype theory. Eleanor Rosch (1978), for example, has suggested that the notion of a prototype helps us to understand vagueness since prototypicality is a property of degree, and vague predicates are associated with such properties. However, it turns out that prototypicality, in this sense, is orthogonal to vagueness, as demonstrated by the fact that an absolutely definite case may have low prototypicality (as penguins do relative to their classification as birds). Indeed, even boundary-drawing concepts induce prototypicality scales. Thus 2 is highly prototypical for *even number*, but it is no surprise to learn that there are plenty of even numbers which have a very low prototypicality rating for this concept: many even numbers are extremely unlikely to be chosen in teaching or exemplifying the concept.

A more promising alternative source of understanding boundarylessness is the parallel distributed processing (PDP) model of the material basis of cognitive activities. Crucial to such a model is another notion of a prototype: an object which has played a causal role as a positive instance in so adjusting the weights of the hidden elements of the network as to help tune it to its recognitional task. This might turn out to correspond more closely to what, a moment ago, I called a paradigm.

Any attempt to describe boundarylessness in such psychological or neurophysiological terms will, however, miss the normative features. We might explain confident application of, say, “red” as the organism’s response to a high level of activation of the output of the red-recognizing network, and a reduction in confidence by a combination of a reduction of the level of this output and an increase in the level of output from the yellow-recognizing network. There may also be explanations in neurophysiological terms of the tendency to include more colours in the reds if you start from reds and move gradually to the yellows than if you reverse the direction of application. But no such facts will begin to capture such aspects of the use of the word “red” as the mandatoriness of its application in some central cases, the freedom available for borderlines, and such rules as that anything at least as red as a red thing is not merely likely to be called “red” but ought to be so called.

The general point is that the vagueness of a vague expression need not, and perhaps should not, feature in a psychological account of how it is used. This account may describe the dispositions-at-a-time to use the expression in terms of a probability function, and may describe a more enduring state in terms of a range of such functions. But the psychologist’s task would be made no harder if he resisted anything homophonic in

describing the inputs to the functions. Thus he might describe the input to a person's function relating to the ascription of "red" not in our colour vocabulary but in terms of the physical constitution of the light striking the eye. So while one should expect harmony between the semantic fact of boundarylessness and such psychological descriptions, the latter can never exhaust the former.

## 9 Conclusion

Let us take stock. Vagueness should be characterized as boundarylessness, not merely in terms of borderlines. Boundarylessness cannot be described sharply, for example set-theoretically; so, whatever insight psychological descriptions may offer, the only semantic description which appears plausible is vague, for example homophonic. We must reject the classical picture of classification by pigeon-holes, and think in other terms: classifying can be, and often is, clustering round paradigms.

Just how widespread vagueness is can be underestimated. Let me draw attention to an area sometimes wrongly thought to be free of it: biological species. Even in quite recent philosophy, there is a tendency to suppose that species come in the "eternal and fixed forms" beloved, according to John Locke, of the Port Royal logicians. It may seem that *strawberry* draws boundaries, since there are no borderline cases. But this is just an accident. There could very well be, and no doubt with the advent of genetic engineering soon will be, a series of plants between strawberries and raspberries, many of them borderline for both concepts. Such concepts do not impose boundaries, but constitute one of the largest and most impressive systems of contrary boundaryless concepts. Locke was right to draw attention to the lack of boundaries by reminding us of boundary-defying "monsters".

One practical application of work on vagueness is in cognitive science, where a possible goal is to implement in machinery the vagueness of our concepts. Another application has already been mentioned. The law must rule a boundary between legitimate and illegitimate acts. Here, boundarylessness would be out of place. Yet such rulings must often traverse territory spanned by a boundaryless concept, like that of being a person. Given the nature of boundarylessness, semantics give freedom. There is some number of minutes such that the nature of the concept of a person, together with the nature of the world, makes it neither mandatory nor impermissible to apply the concept to a foetus of that age in minutes. Hence arguments that use the vague concept to establish or overthrow a sharp ruling are alike inadequate. We can no more argue that aborting a foetus of this age is right because it is not a person than we can argue that it is wrong because it is a person, if *person* is vague at the crucial point. In general, only a pragmatic justification could be found for drawing a legal line in an area where there are no relevant boundaries.

I mention this merely as an example of a possible application whose details remain to be worked out. It proleptically exemplifies my hope that work in the philosophy of vagueness will enable us better to understand how the demands of law and morality should be tailored to the boundaryless fabric of most of our thought and talk.

# Russell on names and communication

---

## I Introduction

A description theory of names claims that all names of some category are related in some way to definite descriptions. Theories may differ about what the relevant category is, and how it should be specified, but let us ignore such differences. More interestingly, theories will differ about what the relevant relation is. For the moment let us abstract from such differences by using the schematic  $R$  to indicate whatever relation is desired.

Description theories differ not only in how they specify  $R$  but also in their structure. Is it claimed that  $R$  holds universally across the language, whatever the speaker, whatever the occasion? Or is the extension of  $R$  speaker-dependent, so that name-description pairs may vary from speaker to speaker? Or is it occasion-dependent, so that even for a single speaker there may be different name-description pairs on different occasions? To allow for these variations it is convenient to think of  $R$  as a four-place relation between a name, a description, a speaker, and an occasion. Then we might represent the three structural possibilities just mentioned as follows, using “ $n$ ” to range over names of the relevant category, “ $d$ ” over descriptions, “ $s$ ” over speakers, and “ $o$ ” over occasions:

- (1)  $\forall n \exists d \forall s \forall o (Rnds o)$
- (2)  $\forall n \forall s \exists d \forall o (Rnds o)$
- (3)  $\forall n \forall s \forall o \exists d (Rnds o)$ .

(1) claims that for every name (of the selected category) there is a description such that for every speaker on every occasion the name is  $R$ -related to the description (for that speaker on that occasion). (2) and (3) specify theories with a progressively weaker structure.

In “Naming and Necessity” (1972), Kripke considers description theories which I think he intends to have the structure of (1): the  $\exists \forall \forall$  structure, for short, for the initial quantifier, being common to all the theories, can be omitted. The hesitation is due to the fact that an official statement of the



structure of the theories he wishes to consider is consistent with, though it does not entail, the  $\forall\exists\forall$  structure of (2).<sup>1</sup> As for *R*, he explicitly distinguishes two possibilities: that *R* is *fixes the meaning* and that *R* is *fixes the reference*. There are therefore up to four theories in play, two alike in structure (that is, in their quantifier prefix) and two alike in the content of *R*.

Kripke on more than one occasion attributes “the” description theory of names to Russell.<sup>2</sup> However, Russell’s theory of names and descriptions has the  $\forall\forall\exists$  structure of (3), and the relation it ascribes is neither of Kripke’s candidates, but rather *makes explicit the thought in the mind of the speaker (on that occasion)*.<sup>3</sup> Hence for two sufficient reasons Russell’s theory is not any of the theories explicitly addressed by Kripke: it differs in structure from any theory discussed by Kripke, and it differs in its specification of the relevant relation between names and descriptions. The following table makes this plain:<sup>4</sup>

	$\exists\forall\forall$	$\forall\exists\forall$	$\forall\forall\exists$
fixes meaning	1	2	7
fixes reference	3	4	8
fixes subject’s thought	5	6	9

The views explicitly addressed by Kripke do not go beyond 1–4. The view I attribute to Russell is 9, which in full is as follows:

**DTN9** – For every name in the relevant category, every speaker, and every occasion of the name’s use, there is a description such that in order to make explicit the thought in the mind of the speaker on that occasion, one must use a sentence in which the name is replaced by the description.

- 1 Thus the numbered theses in Kripke (1972) at p. 280, and again at p. 285, use a schematic “A” for the speaker. One natural way to quantify over the conditions would yield the  $\forall\exists\forall$  structure, but most of Kripke’s discussion seems aimed at the stronger  $\exists\forall\forall$  structure.
- 2 For example, “Frege and Russell certainly seem to have the full blown theory according to which a proper name is not a rigid designator and is synonymous with the description which replaced it” (Kripke 1972: 277); “Frege, Russell, Searle, Strawson and other advocates of the description theory” (Kripke 1972: 349, n. 38). Kripke is not alone in making this attribution to Russell. Michael Dummett describes Russell as holding that “all ordinary proper names are disguised definite descriptions” (Dummett 1973: 97).
- 3 Russell does allow, as we shall see, that a description may sometimes be what one might call reference-fixing; but this is not the relation which is constitutive of his theory of names, and it looks rather different within a genuinely Russellian (as opposed to Kripkean) perspective.
- 4 The table does not do justice to some interrelations between structure and content. Thus, fixing meaning (taken as a feature of the public language) requires the  $\exists\forall\forall$  structure.

In §2 below, I present evidence for regarding Russell as holding DTN9. This theory, as opposed to some lower-numbered description theories, does not have any immediate consequences either for the public reference of a name or for the truth conditions of utterances containing names. In §3, I sketch Russell's account of public reference and in §4 his account of truth conditions. I claim that Russell's theory of the meaning (in more or less Kripke's sense) of *ordinary* proper names is about the same as Kripke's: they are rigid designators whose meaning is their bearer. Sic! In §5 I show how Russell's theories can offer solutions (good, bad, or indifferent) to familiar problems about existence, identity, and belief. In §6, I consider to what extent Russell's theories are vulnerable to attacks on other description theories.

## 2 What Russell actually said

Common words, even proper names, are usually really descriptions.  
(Russell 1912b: 29; 1911b: 156)

This cannot be literally true, and Russell cannot have thought it was, since the categories of names and definite descriptions are syntactically defined, and defined as disjoint (setting aside such atypical expressions as “The Holy Roman Empire”). So the quoted sentence is metaphor or ellipsis. A standard interpretation is that it means that proper names are really *synonymous* with descriptions. One has only to read Russell's immediately subsequent amplification to see that this interpretation cannot be correct:

That is to say, the thought in the mind of a person using a proper name correctly can generally only be expressed explicitly if we replace the proper name by a description. Moreover the description required to express the thought will vary for different people, or for the same person at different times. The only thing constant (so long as the name is rightly used) is the object to which the name applies.<sup>5</sup>

This quotation concludes the exegetical task, for the first sentence establishes my claim about how Russell understood the *R*-relation, and the second establishes the occasion-dependent  $\forall\forall\exists$  structure. However, it will be useful to say something about the nature of the enterprise on which Russell was engaged,

5 The tone of these remarks contrasts strongly with that of comparable passages in Frege (e.g. the footnote about the sense of the name “Aristotle” (1892a), and the discussion of “Dr Lauben” (1918). Frege finds the variation in the descriptions problematic from the point of view of his theory, and is forced to regard it as a lapse from perfection. There is no sign of such an attitude in Russell. The variation is not inconsistent with anything else he is tempted to believe.

and, with this in mind, to comment briefly on some passages which appear to lend greater support than the one just quoted to the interpretation of Russell according to which his *R*-relation is synonymy.

Russell's conception of philosophical problems has a highly individualistic character. The subject and its thoughts and experiences are the relatively unproblematic starting-point. More exactly, to allow for his conversion to the no-self theory of self, of which the beginnings are visible at the time of the above quotation, at the centre are directly accessible thoughts and experiences. Basic thoughts have as their subject-matter these experiences and their properties. What is problematic is to allow for the external world and people as objects of thought and as objects of knowledge.

How does a public natural language, and its associated notion of *meaning*, appear within such a perspective? The fundamental phenomena are individual experience-related thoughts, assumed to be available independently of language.<sup>6</sup> Different subjects will have basic thoughts with different subject-matters. Each will try to build thought and knowledge outwards from himself. But the idea that each would end by constructing a significantly overlapping set of thoughts, apt to be the public meanings of public language sentences, would be gratuitous.<sup>7</sup> Certainly, there could be no identity of singular basic thoughts, since the experiences which form their subject-matters do not overlap. The most optimistic hope is for overlap in our non-basic thoughts, but if one takes Russell's epistemological assumptions with full seriousness, we would have no good reason to believe that there is any such thing, however smooth the processes we call communication appear, however similar the assignments of truth value by different subjects to the same words. A weaker hypothesis, proclaiming merely some kind of harmony between the thoughts different subjects attach to given words, will as well explain the data as the stronger hypothesis of identity.

6 See, for example: "When one person uses a word, he does not mean by it the same thing as another means by it. . . . It would be absolutely fatal if people meant the same thing by their words. It would make all intercourse impossible . . . because the meaning you attach to your words must depend on the nature of the objects you are acquainted with, and . . . different people are acquainted with different objects." (Russell: 1918: 195) "It seemed to me that belief and knowledge have pre-verbal forms, and that they cannot be rightly analysed if this is not realized" (Russell 1959: 132). Moreover, Russell retrospectively held that before 1918 he had not paid much attention to "linguistic" matters. (See Russell 1959: 154: "It was in 1918 . . . that I first became interested in the definition of 'meaning' and in the relation of language to fact".) If this comes as a surprise, it is because one has not done full justice to Russell's idea that the theory of descriptions, and the related doctrines that we assign to his "Philosophy of Language", were for him primarily doctrines about thought rather than language.

7 Consider, for example, the implicature of "If two people are both believing that two and two are four, it is at least theoretically possible that the meanings they attach to the words *two* and *and* and *are* and *four* are the same . . ." (from Russell (1912a), quoted in Russell (1959: 161)). Theoretically possible, since this is a non-basic belief.

Someone who holds this picture would never think of identifying *R* either with *fixes the meaning* (=public language meaning) or with *fixes the reference* (=public language reference). The reason is that one is quite lucky if there turns out to be such a thing as *public* meaning or reference at all, and in any case an adequate theory will need to reconstruct such phenomena from more basic ones. There is no ground-level public meaning or reference, because at ground level all each individual can think about, mean, or refer to is his or her own experience and its properties.

We will see later how Russell can effect some of these constructions. But, in order to complete the exegetical task, I need to consider passages which are apparently in conflict with my interpretation.

- (1)  
 “Apollo” means really “the object having such-and-such properties”, say “the object having the properties enumerated in the Classical Dictionary”.  
 (Russell and Whitehead 1910: 31)

There is no gainsaying that these words have a reading which conflicts with my interpretation, but consider the attached footnote:

The same principle applies to many uses of the proper names of existent objects, e.g. to all uses of proper names for objects known to the speaker only by report, and not by personal acquaintance.

Since two speakers may vary in this respect concerning a name both use in apparently successful communication, the appending of the footnote to the text suggests that the text’s use of “means” is not the public language use we might imagine, but rather something idiolectal.

- (2)  
 When I say, e.g., “Homer existed”, I am meaning by “Homer” some description, say “the author of the Homeric poems”, and I am asserting that those poems were written by one man.  
 (Russell 1918: 252; cf. the discussion of “Romulus” at p. 242.)

First, what is explicitly addressed is (possibly idiosyncratic) speaker meaning, rather than public meaning. Second, the name occurs in an existential sentence and, as we shall see, the position I attribute to Russell has it that these sentences will be accorded rather special treatment.

- (3)  
 The names that we commonly use, like “Socrates”, are really abbreviations for descriptions . . . .  
 (Russell 1918: 200)

However, the passage continues:

We are not acquainted with Socrates and therefore we cannot name him. When we use the word “Socrates”, we are really using a description. Our thought may be rendered by some such phrase as, “The Master of Plato”, or . . . . (Russell 1918: 201)

Here the later amplification is consistent with my proposed interpretation.

(4)

If somebody mentions Socrates and you have never previously heard of him, you can look him up in the encyclopaedia and you may take what you find there as the definition of the word “Socrates”. In that case, “Socrates” is not, for you, strictly a name, but a substitute for a description . . . I do not suggest that in ordinary language or in grammar we should refuse to regard “Socrates” (say) as a name, but, from an epistemological point of view, our knowledge about him is very different from our knowledge of things with which we are acquainted. In fact, everything that we know about Socrates can only be stated fully by substituting some description of him in place of his name, since, for us, it is only from the description that we understand what the word “Socrates” means. (Russell 1959: 168–9)

If the earlier sentence is read in the light of the later ones, my interpretation is, I believe, the natural one. A full statement of an individual’s knowledge is one thing; an account of what is involved in his communications is another.

### **3 Russell’s theory of communication and the public reference of names**

Although Russell takes idiolect meaning as primary, he thinks that one can make derivative sense of the successful use of a name in communication. It is not a matter of thought-sharing, but of reference-sharing.

[W]e often *intend* to make our statement, not in the form involving the description, but about the actual thing described. That is to say, when we say anything about Bismarck, we should like, if we could, to make the judgement which Bismarck alone can make, namely, the judgement of which he himself is a constituent. In this we are necessarily defeated, since the actual Bismarck is unknown to us. But we know that there is an object *B* called Bismarck, and that *B* was an astute diplomatist. We can thus *describe* the proposition we should like to affirm, namely, “*B* was an astute diplomatist”, where *B* is the object

which was Bismarck. What enables us to communicate in spite of the varying descriptions we employ is that we know there is a true proposition concerning the actual Bismarck, and that, however we may vary the description (so long as the description is correct), the proposition described is still the same. This proposition, which is described and is known to be true, is what interests us . . . .

(Russell 1912b: 31; 1911b: 158)

The idea is that one using a sentence containing a name intends that he and his audience will identify (by description) the same singular proposition. In successful communication, this does actually occur (and, perhaps, is known to occur). We could put this “description theory of communication” more explicitly as follows:

**DTC** – A speaker, S, who utters a sentence containing a name  $\alpha$ , and who satisfies DTN9 in virtue of associating “the  $F$ ” on this occasion with  $\alpha$ , communicates successfully to an audience A only if:

- (i) S intends that:
  - (a) A should associate  $\alpha$  with some definite description,  $\delta$ , and
  - (b) something should uniquely satisfy “ $F$ ” and should uniquely satisfy the predicate in  $\delta$ , and
  - (c) A should recognize that S wants A to “be interested in” some singular proposition containing as a constituent an object which each can think of by means of whatever description he associates with  $\alpha$ , and
- (ii) S’s intentions in (i) are fulfilled and
- (iii) S and A know that (ii).<sup>8</sup>

The idea is that this necessary condition for successful communication falls

8 Charles Daniels and Peter Hylton both raised the question how the quotation from Russell, this elaboration of it, and the further theories I am about to elaborate on Russell’s behalf would look if one took seriously Russell’s claim that there are no such things as propositions. There is no complete answer, since the theory of names involves attributing to speakers and hearers general beliefs, and we do not know how Russell’s multiple-relation theory of belief deals with these. But the theories of names raise no special problems. Call an object known only by some description a “pseudo-constituent” of a mental state whose content is expressible by that description. Pseudo-constituents will not themselves be terms of the belief-relation, but they will correspond to groups of such terms in whatever way the multiple-relation theory of belief adopts in analysing generality. (If Russellian logical form pictures the mind, then a belief involving the  $F$  known by description will correspond to something like the sequence:  $\langle \exists x, Fx, \&, \forall y, Fy, \rightarrow, etc. \rangle$ .) Then the main line of Russell’s theory is that objects typically enter only as pseudo-constituents into mental acts of communicating using names, even though they are witnesses to the existentially quantified conditions for such acts to constitute successful communicatings.

short of being sufficient only in so far as it fails to address the role in communication of expressions other than names. We will turn in a moment to what the “interest” specified in (i)(c) is supposed to be, and how it is supposed to relate to the name’s bearer. DTC speaks to the “standard” use of names. As we shall see, an overall view will have to allow for non-standard uses as well.

On this basis, a natural account of public reference would (as a first approximation) take the following form.

**DTR** –  $x$  is the public referent of a name  $\alpha$  in a group  $G$  iff whenever a member of  $G$  uses  $\alpha$ , then he and any audience meet the conditions set by DTC and  $x$  satisfies the quantification in (i)(b); that is  $x$  is the  $F$  and  $x$  is  $\delta$ .<sup>9</sup>

It would no doubt be better to relax the condition so as to allow for occasional failures of successful communication, perhaps by saying that a public referent in  $G$  is also a public referent in any group of which  $G$  is a suitably authoritative and informed subset. Such nuances are more detailed than is appropriate in discussing Russell’s very undetailed suggestions.

Russell, like some subsequent commentators, is sometimes less than strict in his use of the expression “description”, as it occurs in description theories. Officially, “description” abbreviates “definite description”, that is, a phrase of the form “the so-and-so”. However, Russell sometimes appears to use the phrase more loosely, to include information which does not have any uniqueness built into it. Thus, he says in connection with our use of “Bismarck” that “the description in our minds will probably be some more or less vague mass of historical knowledge”, and in connection with my use of “Julius Caesar” that the description will be “made up of some of the things I know about him”. Historical knowledge, and the things I know about someone, do not always or even typically have uniqueness built into their content. We may thus be tempted to contemplate a different theory of public reference in which the information associated with the use of a name may not be unique, but in which uniqueness enters in the following way: the referent of a name is whatever object is uniquely such that no other object satisfies more associated information.

We might also contemplate yet another variant: the public referent of a name is whatever object is uniquely the source of the associated information. Though perfectly consistent with DTN9, we here move away from anything which could properly be attributed to Russell.

What is the “interest” which an audience is supposed to display in some

9 The descriptions “the  $F$ ” and  $\delta$  may, of course, vary from occasion to occasion of the use of  $\alpha$ .

singular proposition? Russell is not completely explicit, but it seems that the audience is intended to regard the speaker as having spoken truly just on condition that this singular proposition is true. Note Russell's emphasis in the previous quotation: neither speaker nor hearer need *know* (that is, understand, that is, be acquainted with) the singular proposition in order for it to play this role in their thoughts. For both, it may be accessible only by such descriptions as: the singular proposition whose first member is the first chancellor of the German Empire and whose second member is the property of being astute.

#### 4 Russell's theory of the truth conditions of utterances containing names

"The only thing constant (so long as the name is rightly used) is the object to which the name applies." The quotation continues:

But so long as this remains constant, the particular description involved usually makes no difference to the truth or falsehood of the proposition in which the name appears. (Russell 1912b: 30; 1911b: 156)

I think we can read Russell as here specifying the truth conditions of utterances containing names, which I will dignify by an acronym:

**DTT** – An utterance of a sentence "...  $\alpha$  ...", where  $x$  is the public referent (in the sense of DTR) of  $\alpha$ , is true iff  $x$  satisfies "...  $\zeta$  ...".

#### Comments

First, the truth condition treats "...  $\alpha$  ..." as expressing a singular proposition, in the sense that an object enters into its truth condition. The truth condition is given by what occurs to the right of "iff", and there is a familiar Russellian contrast between logically proper names, which introduce an object into a truth condition, and definite descriptions, which do not. The truth condition introduced by DTT belongs firmly with the former, rather than the latter, since the bound variable " $x$ ", unlike a descriptive phrase, is assigned an object before the truth condition is evaluated. In my view, the contrast between those grammatically singular expressions which have, and those which lack, object-involving truth conditions is the contrast which underpins Kripke's distinction between rigid and non-rigid designators, and it is available even if our metalanguage is not modal. (For example, even without modality we can distinguish object-involving truth conditions as ones which are absent in the absence of an appropriate object, and as requiring some distinctive epistemic relation between the thinker and the relevant object.) I think one can therefore say, with only



unimportant inaccuracy, that DTT treats *ordinary* proper names as rigid designators.<sup>10</sup>

The descriptions which, by DTN9, are associated with names are not (typically) themselves rigid designators. But (as Kripke would be the first to admit) a non-rigid designator can be used to specify a rigid reference condition, that is, a condition which ensures that an expression which satisfies it is a rigid designator. (Thus we could stipulate that the inventor of the zip is, with respect to every world, what “Julius” is to designate, if that person exists at the world in question. The stipulation would ensure that “Julius” designates, with respect to any world at which it designates anything, the person who in the actual world invented the zip. Yet the description “the inventor of the zip” is non-rigid.)

Second, if we take DTR and DTT together we can justify hearing the following words (as they would typically be heard by our contemporaries) as a true account of Russell’s theory of ordinary proper names: their meaning is their bearer. Contemporary ears are unlikely to demand more of *meaning* than that it be both what is common between speaker and hearer in successful communication and what determines truth conditions. Bismarck himself is, Russell says, both all that is typically in common between a speaker and a hearer in virtue of the speaker’s communicatively successful use of “Bismarck”, and the name’s contribution to truth conditions.

Third, although there are many statements of DTN9 in Russell’s writings, there is scant evidence for DTC, DTR, and DTT: I believe I have quoted almost all of it. This is not surprising in the light of Russell’s generally individualistic stance and his official lack of concern with language before the 1920s.

Fourth, many description theories, for example DTN1, entail (or make almost irresistible) theories of communication, reference, and truth conditions. By contrast, DTN9 leaves these matters open. In particular, it leaves open the possibility that the thought expressed by the speaker – the psychological unit – should be distinct from what the speaker thereby publicly expresses – the communicative unit. This separation ought to be congenial to Kripke (and to writers like Brian Loar 1988).

Fifth, the theories are designed to cater for the case in which it is “very

10 Gary Wedeking made me realize the inaccuracy. Strictly, we can speak of a rigid designator in Kripke’s sense only if we adopt a modal metalanguage, which would be uncongenial to Russell. However, the only appropriate way to set DTT in such a language ensures that it treats the names to which it applies as rigid designators in exactly Kripke’s sense. For the correct modalization of it, in possible worlds terms, is: where  $x$  is the actual public referent (in the sense of DTR) of  $\alpha$ , an actual utterance of a sentence “. . .  $\alpha$  . . .” is such that, for any world,  $w$ , it is true with respect to  $w$  iff  $x$  satisfies “. . .  $\zeta$  . . .” with respect to  $w$ . The (implicit) uniqueness quantifier which binds “ $x$ ” lies outside the scope of the possible worlds quantification.

much a matter of chance which characteristics of a man's appearance will come into a friend's mind when he thinks of him" and thus serve as associated descriptions (Russell 1911: 157; 1912b: 30). This is consistent with there being names for which this is not so: for which there is, instead, a conventionally associated description. For such names, a theory with  $\exists\forall\forall$  structure will be appropriate. Russell suggests a pervasive possibility: the description "the person called 'N'" is available to all "consumers" in a name-using practice.<sup>11</sup>

Lest a name should have some conventionally associated description, the following rule of thumb is to be commended: when reporting a person's thought or speech, by default use the name they would or did use, if it is in your repertoire. If descriptive associations are standard for this name, then you will catch those too (and thus get closest to your subject's thoughts). If they are not, nothing is lost. Of course, the default position will be overridden if you have special knowledge of the particular descriptive associations.

## 5 Solutions to some problems

There are a range of standard semantic problems which description theories are meant to solve. Some might challenge my attribution to Russell of a description theory of names as weak as DTN9 on the grounds that this theory will not seriously engage with, let alone "solve", these problems. In this section, I discuss three of these problems. Typically, it is correct that we will need to look beyond DTN9, and that we will need to modify some of DTC, DTR, and DTT. However, I will maintain that we need not depart from Russell's general ideas.

### **Existence**

One problem concerning existential sentences is to explain how there can be true negative ones, like "There is no such thing as Vulcan". For it seems that the semantic role of "Vulcan" is to refer to something. If it does this, then the sentence will be false; but if it does not do this, it would seem that the sentence ought to be meaningless, through containing an expression which is failing to fill its semantic role. In either case, it would not be true.

But if "Vulcan" were synonymous with a definite description as DTN1 has it, and if one espoused a suitable theory of descriptions, Russell's for example, this problem would disappear. Such a sentence would be synonymous

11 This terminology comes from Evans (1982), who acknowledges Kripke. However, the contrast is already present in Russell, for example in his distinction between two kinds of users of "Bismarck" in the passages adjacent to those already quoted (1911b, 1912b). It occurs elsewhere, including a mention as late as 1959: 168.

with something like “There is no such thing as a unique intra-mercurial planet”. Since this contains no expression whose semantic role appears to be to refer, there is no appearance of a dilemma between meaninglessness and falsehood. This is doubtless what Kripke has in mind when he says that a theory upon which ordinary proper names are synonymous with descriptions can solve problems about existential statements.<sup>12</sup>

Is any such solution made possible by the theories Russell actually holds? Clearly the possibility of true negative existential sentences is *consistent* with DTN9. (DTN9 does not *preclude* the stronger DTN1.) And DTR correctly predicts that the names in such truths will have no public referent. However, the possibility of using them in successful communication is not allowed for by DTC, and their having truth conditions, and so being true, is not allowed for by DTT.

DTN9 allows for the mind of a user of a negative existential sentence to contain a truth, one of the form: there is no unique *F*. One straightforward position for Russell to adopt is that such a sentence cannot be guaranteed to have public truth conditions. From his perspective, this is not very shocking: public truth conditions are in any case a matter of good fortune, representing a particular kind of harmony within a linguistic group.

It will often happen that an empty name will be associated with a much narrower range of descriptions than a non-empty one. (One does not understand “Santa Claus” unless one associates the name with Christmas.) This is because there is no object to give rise to discrete bodies of information. So an alternative to the view that negative existentials have no public truth conditions is that descriptive truth conditions can be correctly assigned when there is a standardly associated description, which there often is for empty names. In this case, the strength of a theory with  $\exists\forall\forall$  structure will be appealing.<sup>13</sup>

But from Russell’s viewpoint, such details matter little. The main point is that the thought in the mind of an utterer of a negative existential has a clear and determinate truth condition, and what we say after that is arbitrarily determined by the work we would like our theory of public reference and truth conditions to do.

It is interesting that Russell, so far as I know, *always* applies his description

12 “If you give up the idea that [the theory] is a theory of meaning and make it into a theory of reference . . ., you give up some of the advantages of the theory. Singular existential statements and identity statements between names need some other analysis” (Kripke 1972: 277).

13 One can’t count on such a conventional association for all empty names. Suppose the original plan is that “Vulcan” should denote a unique planet between Mercury and the sun, but that subsequent theory has it that Vulcan emits rays which are uniquely responsible for cancer. One who utters “Vulcan does not exist” may, on Russell’s account, *think* e.g. that there is no planet between Mercury and the sun, or that cancer is not caused by planetary emissions, but concerning neither thought should he count on his hearer associating it with his words.

theory to negative existential sentences in such a way as to emphasize the idiolectical nature of the truth conditions assigned (*i.e.* truth conditions for the thought in the mind of the speaker, rather than anything more public). The standard pattern has already been quoted (as (2) of §2): “when *you say* N does not exist *you mean* the F does not exist”.

### **Belief**

One problem about belief ascription is this. If “Phosphorus is visible in the morning” can correctly be used to ascribe a belief to the ancient astronomers, why cannot “Hesperus is visible in the morning” also be so used, given that Hesperus is Phosphorus? Yet, intuitively, “The ancient astronomers believed that Hesperus is visible in the morning” is false.

The problem would be solved on DTN1, since the two names would be synonymous with different and non-synonymous descriptions, and it is not surprising that a sentence with one meaning should give a correct attribution of belief and a sentence with another an incorrect one.

Russell has little to say about belief ascriptions; at any rate, little which bears on this problem. But we could interpolate on his behalf the following line of thought. A belief attribution should aim to represent what it is like in the mind of the believer. Hence the standard account of the use of a name is suspended. In this case, we ideally aim to excite in the hearer a thought which is identical to the thought believed. In the example just given, it is obvious why “Phosphorus” comes much closer to that ideal than does “Hesperus”. Indeed, “Phosphorus” is close enough to the ideal for our attribution to be judged true, and “Hesperus” so far from it that our attribution will be judged false. Very likely, there will also be some indeterminate cases.

The rule of thumb mentioned earlier will usefully commend using the same name as that which the believer would or did use.<sup>14</sup>

### **Identity**

One problem about identity is to explain why identity sentences between names, for example “Hesperus is Phosphorus”, should not be knowable *a priori*, if they are true.

On DTN1, the problem presents as little difficulty as the fact that it is true but not knowable *a priori* that the first heavenly body to appear in the evening is the last heavenly body to vanish in the morning.

On Russell’s actual account, this is one more case in which the standard story about the use of names must be suspended. Something more than the

14 Following this rule of thumb will also yield intuitively better results for *saying* as opposed to *believing*.

singular proposition matters to my communicative purposes when I utter an identity sentence using two names. What matters is that I should take you to identify the relevant singular proposition in a way structurally like my own, by using two definite descriptions rather than one. But this is the only difference: I need not care which you use, and no thought-sharing is required for the success of my enterprise.

The rule of thumb mentioned earlier will usefully commend using just the names which come naturally to you.

I do not say that these solutions are adequate; indeed, they are too sketchy to be adequately assessed. But I do say that they are not nugatory. So even one who thought that Russell's theory of names was intended to address them need not find my interpretation incredible.

However, it is not clear to me that we have cast the problems in the correct form for them to count as ones Russell addressed. To take one example: he was certainly deeply concerned with the question how there could be a true negative existential thought (judgement). However, I challenge the reader to find a pre-1920 passage in which he is clearly and explicitly concerned with the problem of how one gives a plausible interpersonal semantics for a shared language containing negative existential sentences some of which need to be counted as true. For Russell in this period, idiolect (or an individual's thought) is paramount, public language something to which he gave little attention.

## 6 Criticisms of description theories

It is quite often objected to description theories that there simply is no description common to all users of a name. Clearly objections of this type are irrelevant to theories of the  $\forall\forall\exists$  structure.

The following passage exemplifies another common type of objection:

It seems to me . . . absurd to suppose that a beginning student of philosophy, who has learned a few things about Aristotle, and his teacher, who knows a great deal, express different propositions when each says "Aristotle was the teacher of Alexander". (Donnellan 1972: 316)

As Donnellan notes, the supposition would not strike Russell this way, if "express different propositions" means "associate with different thoughts". But the supposition *would* strike Russell as absurd if "express different propositions" means "engage in acts aimed at, and whose truth depends upon, different singular propositions". Russell would say that typically there is a common proposition in such a case, albeit no common thought. Donnellan's objection, as stated, thus carries no weight at all.

Many of Kripke's specific objections deal with public reference and truth conditions, and so do not touch DTN9 itself. But I will also consider whether

they touch any of the theories of communication, reference, or truth conditions which I have attributed to Russell.

Kripke says that if *R* is *fixes the reference* then his modal criticism lapses. His reason is that the reference of a rigid designator can be fixed by a non-rigid one. This is a version of what actually obtains in DTT: the entity upon which the truth of a name-containing utterance, standardly used, rigidly depends is fixed by a thought-component which may not be, and typically is not, rigid.

So the only Kripkean objections which are relevant are those which claim that someone might use a name without having any definite description in mind which denotes what the name publicly denotes. There are two cases: (1) those in which the speaker or thinker is simply unable to produce any information which has the form of being uniquely identifying. The Feynman/Gell-Mann examples fall into this category; (2) those in which the speaker or thinker has a definite description in mind, but its denotation diverges from the public reference of the name. The Gödel/Schmidt examples fall into this category. The first point runs directly contrary to DTN9, the second leaves this theory untouched, but threatens the conjunction of DTC, DTR, and DTT.

Kripke writes:

The man in the street . . . may . . . use the name “Feynman”. When asked he will say: well he’s a physicist or something. He may not think that this picks out anyone uniquely. I still think he uses the name “Feynman” as a name for Feynman. (Kripke 1972: 292)

The best response open to Russell, I believe, is to deny the case. Russell in another connection suggests a way in which this could be done.

. . . in order to discover what is actually in my mind when I judge about *Julius Caesar*, we must substitute for the proper name a description made up of some of the things I know about him. (A description which will often serve to express my thought is “the man whose name was *Julius Caesar*”). For whatever else I may have forgotten about him, it is plain that when I mention him I have not forgotten that that was his name.) (Russell 1911b: 160)

Applying the idea to the present case: we can surely credit the speaker with the description “the person (or perhaps: famous physicist) called ‘Feynman’”. This manifests the pull of the social, which Kripke has so usefully stressed: in our use of names we often defer to the use of others. This is conspicuously so in a case like the one Kripke envisages for Feynman. The speaker would have the overriding aim of according his use of the name with that of others. This is what the envisaged description in effect appeals to.

In connection with appeals to this kind of description, Kripke objects that they are circular, and professes amazement that Russell should have made such a suggestion (1972: 284). Amazement would be in order only if someone claimed that every user of the name had to rely upon a description of this kind. Russell could say that this species of description marks a user as a “consumer” in the name-using practice, but that every such practice must also have “producers”. There is no circularity here.

Let us turn to the other species of objection of this general kind: there may be a description in the speaker’s mind which denotes something distinct from the public referent. Hence public reference and truth conditions cannot be determined by associated descriptions in the manner envisaged by DTR and DTT.

Kripke’s example involves our supposing that Schmidt discovered the incompleteness of arithmetic, and that Gödel passed off the proof as his own. If we associate the name “Gödel” with the description “the person who discovered the incompleteness of arithmetic”, then an utterance by us of “Gödel was born in Vienna” would be true iff Schmidt was born in Vienna. Yet intuitively we have referred to Gödel, not Schmidt, and we speak truly just in case Gödel was born in Vienna; Schmidt’s birthplace is irrelevant (cf. Kripke 1972: 294).

Put more in Russell’s terms, his theory has it that the proposition I would like to affirm by uttering “Gödel was born in Vienna”, and upon whose truth value the truth value of my utterance depends, is one whose first constituent is the discoverer of incompleteness, that is, on the present story, Schmidt; whereas intuitively the proposition I would like to affirm is one whose first constituent is Gödel.

Russell cannot very well deny that this description could get associated with the name for one or more speakers on a number of occasions. If such speakers were in a minority, they could be regarded as using “Gödel” incorrectly, that is, not in accordance with a democratized version of DTR (in which only the majority of speaker-hearer members of *G* need satisfy the condition). But it seems conceivable that widespread *misinformation* about some object or person should affect the majority; and in that case the envisaged response would not be available.

I believe that here Kripke is correct in thinking that Russell’s position needs modification. One possibility would be to make more of the pull of the social by requiring among consumers a dominating deference to others’ usage. The description “the person called ‘Gödel’” would *dominate* any other descriptions in the following sense: a speaker’s intentions as specified in DTC would be conditional upon the description he associates with  $\alpha$  not diverging in denotation from the description “the thing called  $\alpha$  in *G*”, where *G*’ is a subset of the speaker’s linguistic group to which he accords special authority: the experts. Typical candidates for expertise would be the producers of the relevant  $\alpha$ -using practice. These may not now constitute the majority of the users of the name.

The idea is not without plausibility. We may use descriptions in identifying roles in connection with our use of names, but always with the proviso that in so doing we are not diverging from the common referent, regarded as set independently of our own usage.

However, it is not my task to assess the validity of this Russellian response, but rather only to note how it ought to appear from the perspective of a correct interpretation of Russell. First, the objection is not to the theory which Russell took to be of fundamental importance, DTN9, but rather to further, distinct theories (DTC, DTR, DTT) to which, I claim, Russell devoted less than a page in the whole corpus of writings in which his “theory of ordinary proper names” is traditionally found. Second, although the details clearly require attention, and will go beyond anything explicit in Russell, it would be hard to find in the standard writings about “the” description theory of names a decisive reason for abandoning the essentials of Russell’s position: the fundamental facts are idiosyncratic facts about thought, captured in DTN9; from these together with ancillary notions one should construct an account of communication which does not necessarily require thought-sharing, together with associated accounts of public reference and public truth conditions.<sup>15</sup>

15 My thanks to Michael Lockwood and Andrew Irvine for helpful comments on an earlier draft.



# How can some thing say something?

---

## I Background

Russell's multiple relation theory of judgement (MRTJ) brings to the fore two related matters, more general than judgement itself, which Russell found perplexing in the first decade of the century and beyond: the nature of complexes and the twofold nature of verbs.<sup>1</sup> The problem of the unity of the proposition, which in turn lies at the heart of the difficulties Russell encountered with MRTJ, are special cases of these more general ones. The general ones arise whether or not we are concerned to find an account of judgement.

Must we regard complexes as something "over and above" their constituents? Russell answered negatively in the case of what he calls aggregates. "Such a whole", he says, "is completely specified when all its simple constituents are specified" (1903: 140). Some wholes do not meet this condition, and are to be called "unities". For example, the unity *A differs from B* cannot be completely specified by its constituents, since these may form simply an aggregate of the terms, A, difference and B, or alternatively the proposition that B differs from A.

In *The Principles of Mathematics*, he claims that "such a whole [sc. a unity] is always a proposition" (1903: 139); in other words, all unities are propositions.<sup>2</sup> If we hold to this, two potential problems fuse into one: how can meanings form any kind of unity? And how can they form the distinctively propositional kind of unity?

However, Russell does not, and should not, hold that all unities are propositions. For example, a fact will count as a unity, by the test of not being exhausted by its components; so, in particular, will Othello's judging that Desdemona loves Cassio.

The other general problem is discussed in terms of the "twofold nature of the verb" (1903: 49): on the one hand it may be a relating relation and, on the other, a relation in itself (1903: 100). "A relation is one thing when it relates,

1 The importance of these two problems was brought home to me by Griffin (1993).

2 This appears inconsistent with his discussion of denoting complexes, which meet the test for being unities rather than aggregates but which are not propositions.

and another when it is merely enumerated as a term in a collection” (1903: 140). When we say that music is the food of love, the verb or relation *love* appears in itself. When we say that Desdemona loves Cassio, *love* appears in such a way as to relate Desdemona and Cassio.

Russell makes plain that the two problems are connected: “Owing to the way in which the verb actually relates the terms of a proposition, every proposition has a unity which renders it distinct from the sum of its constituents” (1903: 52). In itemizing the constituents, the verb or relation appears “in itself” as opposed to “as relating”; so the proposition is more than just its constituents.

Various difficulties supposedly emerge from the phenomena mentioned.

- 1) An adequate account of the phenomena involves contradiction (1903: 48).
- 2) We are at a loss to say what a proposition is.
- 3) Unities which are not aggregates pose a threat to pluralism.
- 4) There’s a special problem about falsehood, quite independently of any theory of judgement: it seems that if, in the unity *Desdemona loves Cassio*, *love* really relates Desdemona and Cassio, then Desdemona loves Cassio.
- 5) There’s a problem for the MRTJ.

Russell’s rather casual remark in “Lectures on the Philosophy of Logical Atomism” that the MRTJ’s treatment of the verb was “a little unduly simple” seems a little unduly disappointing. Had no progress been made in 15 years?

On what would nowadays seem the central topic, the nature of propositions, I think the answer is no.<sup>3</sup> We understand better how Russell failed to address this problem when we see that his primary concern was the consistency of the existence of unities with pluralism (number 3 in the list above). He was less concerned to say what unities are than to show that allowing them was consistent with his overall philosophy, in which pluralism is underpinned by analysis. This is brought out by a comparison of his response to Bradley with his positional statement of the nature of “analytic realism”. He writes:

Mr. Bradley finds an inconsistency in my simultaneous advocacy of a strict pluralism and of “unities which are complex and which cannot be analysed into terms and relations”. It would seem that everything here

3 I do not mean to imply that he had not tried to make progress. But it is not clear that in his modifications of MRTJ, chronicled by Candlish (1996) and by Griffin (1985, 1986), he had grasped that the root of his problems lay not with judgement but with propositions. I am not aware of any evidence for the view that it would have been obvious to Russell that Wittgenstein was right in writing to him in 1913 that the problems with the theory of judgement “can only be removed by a correct theory of propositions” (Wittgenstein 1974: R. 13).

turns upon the sense in which such unities cannot be analysed. What I admit is that no enumeration of their constituents will reconstitute them, since any such enumeration gives us a plurality, not a unity. But I do not admit that they are not composed of their constituents; and what is more to the purpose, I do not admit that their constituents cannot be considered truly unless we remember that they are their constituents.

(Russell 1910: 354)

No hint here of a positive account of what a unity is. The consistency of unities with the overall project is given pride of place in this passage from “Le réalisme analytique”:

Elle [cette philosophie] est analytique, puisqu’elle soutient que l’existence du complexe dépend de l’existence du simple, and non pas *vice versa*, et que le constituant d’un complexe est absolument identique, comme constituant, à ce qu’il est en lui-même quand on ne considère pas ses relations. Cette philosophie est donc une philosophie atomique.

(Russell 1911: 410)

Concern with the admissibility of unities might distract from concern with their nature.

## **2 How to solve the problem: Candlish’s suggestions**

One could think of “the” problem of the unity of the proposition as composed of several related sub-problems:

- (i) how does one distinguish, among collections of meanings, between those which can be arranged so as to say something (e.g. Desdemona, love and Cassio) and those which cannot be so arranged (e.g. Desdemona and Cassio)?
- (ii) given a collection of meanings (e.g. Desdemona, love and Cassio), how does one distinguish between those arrangements of that collection which do say something (e.g. that Desdemona loves Cassio) and those that do not (e.g. that love Desdemona Cassio)?
- (iii) given a collection of meanings which can be arranged so as to say more than one thing (e.g. Desdemona, love and Cassio), how does one distinguish between the things (e.g. between saying that Desdemona loves Cassio and saying that Cassio loves Desdemona)?
- (iv) given a collection of meanings arranged so as to say just one thing, what cements the meanings together in the required way? What is the nature of the further ingredient or entity involved, here referred to as “arrangement”, over and above the meanings themselves?

I think the central puzzle is located in (iv) and that the others serve to illustrate that (iv) is genuinely puzzling.

Candlish (1996), following Russell, gives central place to a special case of (iii), the case I employed in illustrating (iii). He suggests that either of two moves made by Wittgenstein in the *Tractatus* would have resolved this problem. One is to think of propositions as linguistic rather than non-linguistic, for then unity can be acknowledged without automatic creation of the represented fact. The other is to think of propositional signs as themselves facts, so that there is no need for an explicitly represented relation to be both a propositional constituent and the source of propositional unity. I think it is questionable whether either move is necessary or sufficient for a solution.

Consider the first of these suggestions: the unity of a sentence (i.e. of a proposition thought of as something linguistic) can be acknowledged without automatic creation of the represented fact. Perhaps the thought is that one can allow that the sentence "Cassio loves Desdemona" is a unity, in that it says something, without being obliged to say that Cassio loves Desdemona. But if a sentence can say something false, why should not a collection of meanings? Perhaps it is easy to think of a sentence as ordered, and order can play a special role in connection with problem (iii). But if order is allowed in the story, it can also play a special role in ordering the meanings themselves, a role well adapted to solving problem (iii). I can find only one relevant difference between the level of meanings and the level of language. At the former, Russell seems to have been tempted to explain what it is for a collection of meanings to say something by the fact that its verb "really relates" its terms; it is not tempting to explain what it is for a collection of words to say something by the fact that the verb really relates the names or their referents. However, since Russell realized that the temptation had to be resisted, on pain of making falsehood impossible, this does not appear to be a difference which matters.

Merely moving to the level of language does not seem to make a significant difference, let alone to suffice for a solution. It seems clear that the linguistic analogues of the four problems above are to be resolved, to the extent that they are well posed, by grammar and semantics. It is not clear why such theories should not be mirrored as theories about meanings, rather than about the words which mean them. For example, a rule which would contribute to answering the linguistic analogue of question (i) is that an atomic sentence, a species of word collection which says something, consists in an  $n$ -place predicate and  $n$  names, in a certain order. This could be mirrored at the non-linguistic level: an atomic non-linguistic collection of meanings which says something consists in an  $n$ -place property and  $n$  individuals, in a certain order.

Candlish's other Wittgensteinian suggestion is that "propositions are able to represent facts because the propositional signs are themselves facts" (Candlish 1996: 128). If we had a conception of facts which allows for false

facts, or for which we can make a distinction between whether the fact exists and whether it is instantiated, then we make some progress towards a solution.<sup>4</sup> On this view, any appropriately assembled collection of meanings would be a fact. Truth would be a matter of the fact being instantiated; falsehood its not being instantiated. Wittgenstein does indeed have such a conception (standardly translated as “state of affairs”); but it is not to this conception that Candlish draws attention. Rather, what is supposed to do the trick is that the propositional sign itself is a fact.

The fact “*that* ‘a’ stands to ‘b’ in a certain relation says *that* aRb” (*Tractatus* 3.1432). What is this “certain relation”, and how is it expressed? In the standard example, it cannot be the relation of loving, since signs do not love one another. A better candidate would be the relation of flanking an occurrence of “loves”. But now it is mysterious why the *fact* that “Desdemona” and “Cassio” are related by this relation should be a better candidate for meaning than just the sentence “Desdemona loves Cassio”. To put the worry another way, if it is acceptable to introduce this special syntactic relation (flanking “loves”, in the “Desdemona” then “Cassio” order), why would it not be acceptable to introduce an analogous relation at the level of meanings (flanking love, in the Desdemona then Cassio order)? This is just what Othello’s thought does to Desdemona and Cassio: it places them in the relation of “flanking” love, i.e. of being thought by Othello to be love-related (in the Desdemona then Cassio order). So it doesn’t seem to me that shifting to regarding the propositional sign as a fact is sufficient to resolve the problem.

I have not explicitly addressed the question of whether either moving to language, or moving to regarding the propositional sign as a fact is necessary for a solution. My view is that neither move is necessary. This will emerge in the light of what I think is required.

### 3 Another approach

Are the problems of the unity of the proposition special to Russell’s philosophy, or are they still visible from our contemporary perspective? If so, have they been solved, or simply ignored? I think that the problems remain visible, but they are not often explicitly addressed, despite the fact that, or perhaps because, an adequate solution is available within contemporary received wisdom.

Given the amount of criticism Russell’s MRTJ has received,<sup>5</sup> it is surprising to find apparently similar theories being advanced, or at least taken

4 We do not get all the way, since there is still a question about what makes for the difference between a fact (e.g. that Desdemona loves Cassio) and a collection of meanings (e.g. Desdemona, love and Cassio).

5 Not just by my fellow symposiast (Candlish 1996) but by Wittgenstein (as chronicled in Griffin 1985), Geach (1957: 50), Mackie (1973: 28) and many others.

seriously, by many influential contemporary writers (for example David Kaplan<sup>6</sup>), who do not indicate that such theories bring into prominence any problem of propositional unity. Admittedly, contemporary theorists unify the meanings as a sequence, rather than taking them individually. This removes certain difficulties: it enables the belief relation to be invariably dyadic, rather than having to have variable adicity; and it addresses problem (iii), the problem of distinguishing the different things which could be said by a collection of meanings as a function of their order. Russell could not have tolerated sequences in a complete analysis, consistently with the no-class theory of classes. But it would be a mistake to suppose that allowing sequences would have resolved his problems. Merely ordering meanings, without further devices, cannot be guaranteed to resolve more than special cases of problem (iii), leaving untouched some issues to do with scope;<sup>7</sup> and it does not so much as address problems (i), (ii) and (iv). Do those who are happy to roll out accounts of judgement in which people are related to sequences of entities have up their sleeves an answer to these problems?

Perhaps some look to the Fregean notion of functional application. However, there is no solution in this quarter. The question of what makes the difference between a collection consisting of a function and its potential arguments, on the one hand, and the “insertion” of these arguments into the function, and their insertion in one rather than another order, is of essentially the same kind as our original question. Argument-function unity is of a piece with propositional unity.<sup>8</sup>

One current orthodoxy is that a proposition can be thought of as a set of possible worlds. In its more plausible and cautious form, the set of worlds is not identified with the proposition, but simply specifies truth and identity conditions: a proposition  $p$  is true iff the actual world is included in  $p$ 's associated set, and propositions are identical iff associated with the same set. This theory faces a problem analogous to that of unity: what is the difference between the proposition and the associated set? A proposition manages to say something, to have truth conditions; the set is some kind of model of these. A set cannot be identified with a truth condition, since a

6 E.g. Kaplan (1977). The recent symposium between François Récanati (1995) and Mark Crimmins (held just a week before the Southampton conference at which this paper was delivered) takes a version of MRTJ seriously without manifesting any sense that it raises a problem of propositional unity.

7 For example, the ambiguity in “Harry is a dirty window cleaner” is not resolved by linking the meaning to the sequence <Harry, dirty, window, cleaner>, as opposed to some other sequence each of whose members is one of these elements. More complex set-theoretic constructions, sequences with sequences as members, arguably could resolve all these ambiguities.

8 Only an erroneous interpretation of Frege (in my opinion) would attribute to him an attempt to explain (as opposed to label) this unity in terms of unsaturatedness.

condition, unlike a set, is something which can be satisfied (met, fulfilled) or not.

On another version of the possible worlds theory, propositions are simply identical with sets of worlds. This view confronts another problem resembling that of unity: what is it to employ the contemplation of a set of worlds to entertain a thought, rather than simply contemplating it? If we could answer this question, we could solve the problems of unity. This is not the only way in which, in principle, the problems of unity could be solved, since we might also try to solve them by focusing on some non-set-theoretic mode of combination. This alternative approach would claim that to contemplate anything thus combined is *eo ipso* to entertain a proposition, rather than merely to contemplate some collection of entities.

My view is that a problem deserving the name of the unity of the proposition remains for many philosophers; all those, at least, who place sets of worlds at the centre of their semantic theorizing. But not all philosophers do this, and not all face a unity problem. The problem is absent from, for example, a Davidsonian approach to meaning. This is an account of the meaning of sentences which dispenses with meanings as entities, although we will see that this feature is inessential to the approach's capacity to solve the problems of unity.

In a Davidsonian theory, concatenation of the relevant kind is by definition a way of arranging expressions so that the result has a truth condition; which truth condition depends upon the words concatenated and their mode of concatenation. To contemplate an appropriate concatenation of words with understanding is to appreciate its truth condition. There is no unanswered question about how the sentence manages to say something.

The account would not be satisfying unless the way in which concatenation achieves a truth condition were spelled out. In Davidson's approach, this is achieved by a recursive specification of truth conditions. Names are given reference clauses, predicates satisfaction clauses, a general account is given of how names and predicates combine, and in the light of these one can deduce not only that the result of concatenating an  $n$ -ary predicate with  $n$  names says something, but what it says.

In Davidson's hands, the approach assumes that we are concerned with language, and that meanings as entities are not required. This feature is inessential. One could borrow the recursive approach in order to specify special truth-condition-conferring ways of concatenating Russellian meanings; in doing this one would solve the problems of the unity of the proposition in more or less the terms in which Russell stated them.

Let us use curly brackets to indicate the truth-condition-conferring mode of concatenating meanings (regarded as non-linguistic entities). An expression like “{Desdemona, love, Cassio}” will refer to the result of concatenating the meanings in the list in the special way. The theory will say that this result is true iff Desdemona loves Cassio. In general, for any  $n$  objects,  $o_1 \dots o_n$ ,

and any  $n$ -ary universal,  $\varphi^n$ ,  $\{\varphi^n, o_1 \dots o_n\}$  is true iff  $o_1 \dots o_n$  are  $\varphi^n$ -related.\* Providing the truth condition displays the cement, as demanded by problem (iv): the cement consists in the possession of a truth condition, where this is systematically specified. This also resolves which thing is said, as demanded by problem (iii). Problems (i) and (ii) are resolved (for the atomic cases) by dividing the world into individuals and universals, and subdividing universals according to their degree.

I conclude with five observations.

First, the approach is firmly non-reductive. What it is for a concatenation of meanings to say something is explained by the systematic provision of the saying in question. What else could one expect? In Russell's terms, it might be said that we are treating the relevant kind of concatenation as primitive and undefinable. It is not a relation which exists anyhow, ready to be appealed to by the theorist of propositions or judgements. In this sense, it is *sui generis*.

Russell was quite clear, in the unpublished paper "On Functions" (cited in Griffin 1993), about some necessary conditions for solving the problems: we need to find a distinctive mode of combination (my "concatenation"), which, together with the constituents, determines the complex without itself being a constituent; yet the mode of combination must also be capable, on other occasions, of being a constituent of complexes; on such occasions, it will not be exercising its unifying role. If Russell had borne these points firmly in mind in 1913, his attempt at that time to make use of the notion of logical form in the MRTJ might have taken a different, and Davidsonian, turn: rather than trying to make logical form a constituent of what is judged, each logical form should be seen as one way of concatenating meanings so that something is said. It is the systematic, recursive, progress through the totality of logical forms that makes the Davidsonian account possible.

Second, my suggested approach returns us sharply to Russell's problem of the dual nature of the verb. A constituent of a concatenation is a universal. We need to extract from it something more relational for the truth condition. I did this by keeping " $\varphi^n$ " unequivocal, and tacking on "-related" to reveal its role in the truth condition. This just is the shift from relation in itself to relation as really relating. In this setting, "really relating" can unproblematically be understood in the way that Russell feared would lead to objective falsehoods, since the real relating features only on one side of a biconditional. As we might express the unity: the meaning complex  $\{\text{Desdemona, love Cassio}\}$  is true iff *love* really relates Desdemona to Cassio.

It may well be an essential feature of this approach that full homophony

\* [Added note] *Red* is a unary universal, *love* is a binary universal. The degree of a universal numbers how many terms must be added in order to attain a proposition.



cannot be achieved, though we can approach it more closely by dividing classes of concatenations more finely. Thus for unary atoms, we can say  $\{a, \text{the property of being } F\}$  is true iff  $a$  has the property of being  $F$ . However, I think homophony may well be not completely attainable, on account of the following tendency. If you think of a proposition as a collection of meanings, and think of meanings as, in the first instance, individuals and properties, then it is hard to resist the thought that a unary atom, for example, is most properly described as attributing a property to an individual. The atom is apparently unary in nature, and not just in name; the truth condition binary (involving an individual, a property, and the attribution relation between them). So there's a tendency to see an extra argument place in every proposition. Although this "extra place" conception is, in my view, incorrect, it does not threaten to generate Bradley's regress.\*

The third matter is a question: would this apparatus entitle us to Russell's MRTJ? Russell wanted a multiple relation theory because he thought that if judgement related one who judged falsely to a single thing, it would have to be an unpalatable "objective falsehood". On my proposal, there is no such problem, so the most obvious theory of judgement would relate thinkers to concatenations of meanings. However, there might be other reasons for preferring a theory upon which the mind is related simply to the constituents of the concatenations, where to think of these constituents in a certain way is to concatenate them. Given that we can say recursively what it is to concatenate, this need not be regarded as a mysterious mental power. As far as I know, provided we are happy with multigrade relations, there is no obstacle to this development of the theory I offer Russell.

Fourth, I have considered only atomic concatenations of meanings. For those who, like Wittgenstein and Russell at some periods, think that the logical constants do not denote meanings, it will not be obvious how to extend this approach to non-atomic cases. On this view of the logical constants, the very form of the problems would have to be different, since the constants would supply no meanings (regarded as entities) to be concatenated.

Fifth, and finally, the present concerns raise a further question, not yet mentioned. How can mere lifeless words or meanings (understood in Russell's way, as ordinary individuals and properties), however well selected and arranged, say anything at all? If "mere" is supposed to make us focus on the intrinsic and non-relational properties of words or meanings, then they

\* [Added note] This is because at this stage in the dialectic we are assuming that concatenation is a genuine possibility, so that individual, property and attribution-relation can concatenate to say that the individual possesses the property. Bradley's regress arises when one finds no room for concatenation, finding instead, at each stage of the regress, only further entities requiring to be concatenated.

cannot say anything. They can say something only in virtue of their relational properties, their use. It would be another project to consider the extent to which Russell and Wittgenstein's difficulties about the nature of propositions in the early part of the century can be traced to their not in that period finding room for this crucial notion.

# Easy possibilities

---

## I Introduction

If you know, you could not easily have been wrong. Call this the Reliability Conditional. Its modality relates to the knowing, rather than to what is known. A true proposition which might easily have been false can be known, and a true proposition which could not easily have been false can be believed without being known. If a machine functioned reliably (on a given occasion), it couldn't easily have gone wrong (on that occasion). Reliability and easy possibility are closely related, so an account of knowledge in which reliability plays an important part is an account that connects closely with the notion of what could easily happen.

This note explores whether the Reliability Conditional is consistent with claims about knowledge that Williamson (1994a) has made in elaborating his epistemic theory of vagueness. Having argued that all intelligible expressions draw sharp boundaries, he suggests that we can explain ignorance of borderline cases for vague predicates in terms of “small differences of meaning”. The essence of the explanation transposes to concepts. Suppose some borderline object is red and I believe this. The envisaged explanation of why this belief does not amount to knowledge is that I could easily have been wrong because I could easily have had a slightly different concept of *red*, one which ruled the object not red. I could easily have had a different concept because my practice (which is at least in part fixed by that of my conceptual community) could easily have been slightly different.

On the other hand, Williamson wants to hold that I do have knowledge of the meaning of my vague terms. Suppose that this knowledge can be properly expressed in this kind of way: “red” means *red*. Suppose also that I could easily have had a concept different from *red*, say *red\**. Doesn't this mean that I could easily have been wrong in believing that “red” means *red*, for I could easily have been employing the maverick concept *red\**, in which case I would have falsely believed that “red” means *red\**?

An affirmative answer appears tempting; and, if correct, it would testify to

an incoherence in Williamson's position. However, I shall argue that this appearance is delusory: the notion of what is "easily possible", as applied to knowledge through the Reliability Conditional, turns out to comport well with various demands made by Williamson's form of epistemic theory, in particular with the combination of knowledge claims described in the previous two paragraphs.

## 2 Tracking

The Reliability Conditional is that if you know, you could not easily have been wrong. No equivalence is claimed: the absence of easy possibility is only claimed to be a necessary condition for knowledge, and I bracket all questions about sufficient conditions.

Is "you could not easily have been wrong" equivalent to "you tracked the truth"? Nozick characterizes tracking in terms of counterfactuals: you track the truth of *p* iff both (a) and (b) hold:

- (a) if it were not the case that *p*, then you would not believe that *p*.
- (b) if it were the case that *p*, then you would believe that *p*.<sup>1</sup>

That you could not easily have been wrong does not entail that you track the truth. You believe that you are not a brain in a vat. You could not easily have been wrong: to have been wrong, you would have had to be a brain in a vat, and it seems that this is not easily possible. So the Reliability Conditional seems not to preclude your current belief that you are not a brain in a vat from counting as knowledge. But, at least according to Nozick, this belief does not track the truth, for it does not satisfy the condition that if it were not the case you would not believe it. On the contrary, if it were not the case that you are not a brain in a vat, that is, if you were a brain in a vat, you would still continue to believe that you are not (setting aside doubts posed by semantic externalism).

Suppose I know that *p* and believe that I do not mistakenly believe that *p*. If I were to mistakenly believe that *p*, I would not believe that I mistakenly believed that *p*, so tracking would preclude my knowing that I do not mistakenly believe that *p*, even when it allows that I know that *p*. On the other hand, under the envisaged conditions, I could not easily be wrong in believing that I do not mistakenly believe that *p*; so the Reliability Conditional does not preclude my knowing this.<sup>2</sup>

If you might have been wrong, it is because it might have been that your

1 Cf. Nozick (1981: 172, 176). This is not his finished account of knowledge, which essentially involves reference to methods of belief formation.

2 Cf. Hughes (1996: 319) and Sosa (1996: 276).

belief is false. In this context, “your belief” does not designate rigidly. One possibility is that your belief retains its content but is false. Perhaps (a) addresses this case. Another possibility is that you believe something else, and that other content is false. The relation between this case and tracking condition (b) is not entirely clear. The simplest thought is provided by familiar semantics for counterfactuals, which makes the truth of both antecedent and consequent sufficient for truth. Under this ruling, (a) and (b) can hold even when you could easily have been wrong. Suppose you come to believe a mathematical truth in some random way (for example, you simply guess at the sum of two large numbers, and by good fortune you are right). It is clear that you could easily have been wrong, that is, you could easily have guessed otherwise. Yet it appears that (a) is true in some trivial way, and (b) holds under the familiar semantics.

If one abandons the familiar semantics for counterfactuals, one may still detect some discrepancy between tracking and not being easily able to be wrong. Presumably (b) fails if, were it the case that  $p$ , you would simply fail to believe it, which is not a case of being wrong, but a case of being ignorant. So if we can make sense of your truly believing that  $p$ , yet it not being the case that were  $p$  true you would believe it, we can make sense of this in a situation in which the latter holds because of your failing to believe it rather than your believing anything false; arguably, this is consistent with your actual belief being one concerning which you could not easily have been wrong. For example: I believe that Mary is married because I notice her wedding ring. She is married, but she hardly ever wears the ring, and indeed didn’t intend to wear it on this occasion (she left it on by accident). I have no curiosity about her marital status, so if I had not noticed the ring, I would have formed no belief about whether or not she was married. Arguably, this falsifies (b),<sup>3</sup> so that I don’t track the truth. Yet, arguably, I could not easily have been wrong (given the uniformly prevailing convention in my culture that only married persons wear such a ring). The Reliability Conditional does better than tracking: I do know that I am not a brain in a vat and that Mary is married and that I am not mistakenly believing that  $p$ ; and I do not know the guessed mathematical truth.

The problem raised by the case of Mary is widespread and general.<sup>4</sup> It can be just an accident, a piece of good or bad fortune, that I come to know that  $p$ . I happen to know that Mary is here, because I happened to turn my head (despite being deeply engrossed) just as she came in; so I saw her, but I

3 It certainly does so if a necessary condition for the truth of a counterfactual is: “there be no sufficiently close possible world in which the counterfactual has a true antecedent and a false consequent” (Plantinga 1997: 144). Given Mary’s disinclination to wear her ring, a world in which she is married but I don’t believe it (because my only clue is absent) is close.

4 Thanks to Ernest Sosa for helping me get clearer about this.

might easily not have done. Much of our knowledge is in this way fortuitous. Further, I can know that  $p$  even if it is just an accident that  $p$  is true, and much of our knowledge is of accidental truths. Both kinds of happenstance are consistent with knowledge; the tracking condition, but not the Reliability Conditional, is apt to rule them out. So Nozick was clearly right to think that one could not give an account of knowledge just in terms of tracking. Introducing the further notion of *methods* has to be done with care. The method by which I come to know that Mary is present is perceptual and it leads to knowledge. The method by which someone whose environment contains many fake barns comes to believe truly that a (real) barn is present is also perceptual, yet it does not lead to knowledge. We must say they use different methods, one reliable the other not. Perhaps the Mary-spotter uses the method: infer from Mary-appearances to the presence of Mary. This is a reliable method in the Mary-believer's circumstances. Perhaps the barn-spotter uses the method: infer from barn-appearances to the presence of a barn. This is an unreliable method in the barn-believer's circumstances.

The Reliability Conditional takes the cases in its stride, precluding knowledge of the presence of the barn (because I could easily have been wrong), not precluding knowledge of the presence of Mary (because I could not easily have been wrong). Explaining why it takes them in its stride may well involve reference to methods. But methods might be better thought of as an explanation of the truth of the Reliability Conditional than as a corrective to tracking.

### 3 KK

Arguably, there are cases in which it could easily have been that it could easily have been that the machine went wrong, even though it could not easily have been that the machine went wrong. This would suggest that the operator "it could not easily have been that it is not the case that" does not iterate; so the modality is not one apt to support the KK principle (if you know, you know that you know). This machine has just functioned reliably: it worked correctly and couldn't easily have gone wrong. But when the engineer was making it, he had a choice between installing one of two kinds of switch, and he had no good reason to choose one kind rather than another. As good fortune had it, he chose the one that performed well under the machine's operating conditions. But he could easily have chosen the other one (his choice was arbitrary – perhaps he flipped a coin), and the other switch would have made the machine unreliable. As things are, it is not easily possible for the machine to have gone wrong; uncontroversially (if the story is accepted) it could easily have been that it was easily possible for it to go wrong. Arguably, the story also sustains the counterexample to iteration: on that occasion it was not easily possible for the machine to have

gone wrong, but on that occasion it was easily possible for it to have been easily possible for the machine to have gone wrong.<sup>5</sup>

Reliability has a feature structurally analogous to a “margin for error” principle. A reliable machine will still produce the appropriate output even if the operating conditions had been slightly different (in likely ways). Suppose  $x$  and  $x'$  are two similar inputs, with  $x$  closer than  $x'$  to what is a paradigmatically “normal” operating condition. Suppose that the machine responds reliably to  $x$ . Then it produces the correct or designed output and could not easily have gone wrong. So it would have produced the correct output in response to  $x'$ . But does it follow that it would *reliably* have produced the correct output in response to  $x'$ ? We had best not say that this follows (though of course it may also be true in many cases), for then the machine’s reliability would extend endlessly outwards from the paradigm cases of normal operation in a sorites-like manner: we would be forced to say that it is reliable under all conditions, since any can be reached by a series of small steps from the paradigm ones in which it is, by hypothesis, reliable. When a machine operates well at the outer limits of its range, it may manifest its reliability, that is, its reliable operation in more central cases, without the near-the-limit operations being actual instances of reliable operation.

One moral is that reliability is a relativized concept: good operation is not required under every possible circumstance, but only under some contextually specified range of circumstances. Indeed, it is hard to make sense of the idea of a machine which can operate reliably under *all possible* circumstances, since these include ones with, for example, very different laws of nature, and ones with all sorts of freak demonic interference.

#### 4 Hidden snags

Suppose a mechanism,  $M_1$ , is reliable within the values 4–7 of some parameter. Idealizing, this may mean that it delivers the right output when the values are 3–8; but is not reliable for the values 3 and 8 themselves, since it produces a wrong output for values 2 and below and for values 9 and above. Suppose now that  $M_1$  is coupled to  $M_2$ , which reliably detects the values of the parameter in question, and allows  $M_1$  to operate only if they fall within the range for which  $M_1$  produces the right output (i.e. 3–8). The coupled mechanism,  $M_1 + M_2$ , is reliable even for values 3 and 8, for if a value marginally different from these (in the unfavourable direction) had been realized, there would have been, not an incorrect output, but no output (or, if you prefer, an output

5 As Christopher Hughes pointed out, the interaction between tense and modality leads to many complications. The alleged counterexample cannot be regarded as definitive.

correctly saying “ $M_1$  cannot reliably handle this situation”). Coupling one mechanism with another which is sensitive to the first’s limitations can extend its reliability.

The reliability of human cognitive mechanisms can likewise be extended by an appreciation of their limitations. We are often in states of the following kind: if a way of arriving at a belief had been apt to produce error in the actual circumstances of a specific use, we would have known this (at least tacitly: the belief-forming mechanism in question would be inhibited). Features of our situation (following Sosa, we may call them “hidden snags”) which make us not in this state may defeat knowledge, by shrinking (sometimes devastatingly) the range of our reliability. This explains why many people feel that the barn-spotter does not know that there is a barn before him. The local fake barns are hidden snags, depriving a normally reliable way of telling whether there’s a barn before one (cursory visual inspection) of its reliability (while leaving in place the reliability of other more close-up and searching ways of spotting barns).

Those who say that hidden snags need to be actual and not merely possible if they are to defeat knowledge are certainly right within a reliabilist perspective. Although the perspective suggests that, typically, knowledge requires absence of hidden snags, it does not suggest that it requires knowledge of the absence of hidden snags.

## 5 Determinism

“Could easily” does not need to invoke indeterministic notions. A determinist can properly affirm: “You should not have driven so fast – you could easily have crashed”. The claim could be defended in terms of small differences in relevant parameters, for example: “You would have crashed if the road had been a little more slippery”. There is no need to suppose that the slipperiness of the road is other than completely determined by previous circumstances, or that there is a causal possibility that the actual slipperiness could (with no changes elsewhere) have produced the crash. The point, rather, seems to be that if the road had been a little more slippery, you would not have known. You were rash because you engaged in a course of action which, for all you could know in advance, would have led to disaster.

This example may suggest that “could easily” is consistent with determinism only when the possibility itself issues from ignorance. The impression might be reinforced by the following contrast. The bridge will break under a load of 80 tons but not under a load of 79 tons. I drove my 79-ton truck across and got safely to the other side. Safely? Case 1: my weigh bridge registered 79 tons but is accurate only plus or minus 5 per cent. Then it was rash of me to cross because my truck could easily have weighed enough to break the bridge, so the bridge could easily have broken. Case 2: my weigh bridge was dead accurate, and I really knew all the facts. Then my crossing was not



rash: the bridge could not easily have broken. The second case involves the same small differences as the first, but in this case they are discriminated by the subject. This may suggest that easy possibilities are to be explained in terms of small differences which are significant for the outcome but which could not easily be known.

This impression is just an artifact of the examples so far. If we turn to mere mechanisms, we may reasonably expect a more general feature from which the appeal to knowledge emerges in special (e.g. action-involving) cases. Moreover, the relevant feature is consistent with determinism. A machine is reliable if it is so structured as to produce a good frequency of successful outcomes under a reasonably wide range of conditions (a range determined contextually). In particular, a small variation in input conditions should not normally be enough to tip it into malfunction. The action-involving cases in which ignorance is crucial are just special cases of this. You were lucky to get round the corner without a crash: if you go on behaving like that you'll kill yourself. We're cashiering you for crossing the bridge with that load: a rash fellow like you will undoubtedly cause a disaster sooner or later. The system consisting of you and your vehicle is not reliable, because there are closely similar circumstances under which the desired outcome (getting round the corner without crashing, getting across the bridge) will not be held causally in place by the system's structure. A determinist need find nothing amiss, since we do not have to envisage alleged possibilities in which the inputs are just the same and the outputs different.

## 6 Worlds

It is easily possible for me to be wrong in believing that  $p$  (even if it is true that  $p$ ) iff at some world "close" to the actual world the actual episode of forming the belief that  $p$  (or the counterpart of this episode) is one in which a false belief is formed. In some cases, this is because the same belief is formed, but is false at the close alternative world. In other cases, it is because a different, and false, belief is formed at the close world.

Our notion of a "close" world is not an input to our understanding of knowledge, reliability or easy possibility. Rather, we must fashion the notion of closeness to do proper justice to these notions. For example, no intuitions about closeness should be allowed to rule either that the demon's world is close or that it is not: argument is needed, and the concept of closeness must be so developed as to reflect the outcome.

The initial idea is that worlds are close when there is only a small difference in the value of some parameter. It is because 79 is close to 80 that it was dangerous to cross the bridge with a 79-ton truck. But not just any parameter is relevant. No doubt two worlds that are now completely different may have been close at the time of their Big Bang, for a small difference in parameter values then can lead to endlessly large differences later. When assessing the

reliability of machinery, we have already used the notion of intended operating conditions and intended function. A normal computer is not unreliable because it will not operate under 50 metres of water, but a Rolex which failed to operate under these conditions would have to be returned as defective. We need some notion like that of intended function to specify what counts as correct operation. A screwdriver is not an unreliable screwdriver just because it breaks when used as a lever.

A reliabilist conception of knowledge requires it to be reliably delivered by a mechanism whose job it is to produce true belief.<sup>6</sup> It will accordingly be enough for a difference to count as small that it would not produce any significant difference in the mechanism's operation. Some differences in an aspect of the agent's situation which he would not have noticed count as not having affected his belief-forming mechanism, and so count as small, even if relevant to the truth or falsehood of the belief he actually formed. To use one of Williamson's examples: I believe truly that there are just so many people in a crowded stadium, but I could easily have been wrong because there could easily have been a person more or less: that is, a person more or less would not have affected the mechanisms which issued in that belief, and the same belief would have ensued. So a world in which there is a person more or less and I form the same belief is close. We must be careful to avoid rash generalization, or we will find ourselves counting the difference between our world and a demon world as small, on the grounds that the subject would not have noticed the difference. The talk of "noticing" is misleading, as it suggests a comparison by the subject, whereas what should properly be at issue is whether there would be a significant difference in the subject's overall behaviour. In difficult cases, it will be impossible to assess this other than from the perspective of a theory. (Would my "behaviour" have been "significantly different" in a demon world?) A separate issue is that we may apply different standards in different contexts: a reliable Skoda may break down more often than a reliable Honda, and the demands we place on whether a witness "really knows" may be related to the severity of the crime.

Even given a suitable notion of closeness, we cannot gloss "could not easily have been otherwise" thus: for every world in which it is otherwise, there is a closer one in which it is not. Suppose machines of a certain kind are reliable only if they will work within the range 0–10 of some parameter. This machine has just functioned correctly, with the parameter value 6. However, it will go wrong at values 8 or higher. That is, it could easily go wrong (within its intended operating range) so it is not reliable. Yet for every world in which it goes wrong (with the parameter values 8 or higher) there is a more similar one (because closer to 6) at which it goes right.

<sup>6</sup> This is to be distinguished from "requires it to be delivered by a reliable mechanism". A reliable mechanism may not operate reliably at the outer limit of its normal range.

## 7 Vagueness

That I do not know of a borderline case that it is red, even if I truly believe it, is overdetermined: I could easily have formed a false belief about a different object, for example if an indistinguishable or barely distinguishable non-red shade had occupied my attention at that moment; I could easily have come to the opposite, and so false, view about this very object; and, finally, the possibility Williamson stresses: I could easily have had a slightly different concept, one which, drawing the borderline in a slightly different place, rules the object not to fall under the concept.

This third possibility requires a little elaboration if we are to grasp its knowledge-defeating character. Suppose I stipulate that, in a certain examination, students are to count as “excellent” just on condition that they attain a mark of 72 or more. I could easily have made a different stipulation, for example, that students are to count as “excellent” just on condition that they attain a mark of 73 or more. Does it follow that, when I apply the epithet “excellent” to a student who has scored 72, I do not know that the student counts as excellent, because I might easily have used a concept which, had it played the relevant part in my belief, would have ensured that the belief was false? Obviously not. Knowledge is defeated if the shift of concept functions as a hidden snag, making a way of arriving at a belief we have actually used unreliable without our knowing this (or without inhibiting belief-formation by that method). If we had made a different stipulation governing the local use of “excellent”, setting the minimum at 73, we would know this, and this knowledge would have inhibited the passage from 72 to excellent; so our actual method of forming the judgement, by noting that the student scored 72 or more, is reliable. By contrast, it is possible for the extension of *red* in the thoughts of ourselves and our community to differ from its actual extension in a way I do not detect. This possibility renders unreliable any actual method of reaching the belief that something close to the borderline is red. The first difference changes my belief-forming behaviour, whereas the second does not.

The possibility in question does not attach to every concept. Consider one which is sharp (or let us so suppose), like being a foot long. It would not be easily possible for me to use, in place of *foot*, the concept *metric foot*, without this bringing with it many related differences, despite the fact that there is only a small difference between a foot and a metric foot.<sup>7</sup> If I had the *metric foot* concept in place of that of *foot*, my measuring practices would be different (for example, I would happily measure with a centimetre rule, whereas I actually find the conversion from centimetres to feet too hard); I would associate the price of timber with different numbers, depending as I took the

7 One metric foot is 300mm; a foot is about 305mm.

price to be per foot or per metric foot; in many circumstances, judgements I would actually formulate with the words “that is a foot long” would not be made; and so on. So an episode of coming to believe truly that something is four feet long could not easily have been an episode of coming to believe falsely that it is four metric feet long. Knowledge using non-vague concepts is not impaired. Likewise, knowledge applying vague concepts to clear cases is not impaired. A difference in my concept *red* so great that a clear case of red no longer falls under the different concept is a difference that could not easily occur, since its occurrence would bring with it large scale changes in my behaviour.

Must an epistemic theorist insist on the third source of ignorance? I presume that the motivation is that not all true belief involving vagueness need be perceptual. Suppose that 34 grains (of a given kind) are the fewest that can make a heap. As Williamson stresses, if this is true it is necessarily true. Perhaps someone comes to believe it not on any perceptual basis, but in a way that strikes him as the same as the way in which he comes to know truths of reason. How does the epistemic theorist show that, despite truth and this apparently favourable aetiology, the belief is not knowledge? Appeal to less than perfect perceptual powers seems irrelevant. Appeal to a potential conceptual instability will cover any case, perceptual or not, of purported knowledge in which a vague concept is applied to a borderline case.<sup>8</sup>

Will it cover too many cases, including knowledge of meaning? Earlier I envisaged the claim that one does not know that “red” means *red* because one could easily have had a different concept, *red\**, in which case one would have judged falsely that “red” means *red\**. There seems also to be another instability: the word “red” could easily have meant something different, e.g. *red\**; if it had, then my judgement that “red” means *red* would have been false since “red” would mean *red\**. Why do these possibilities not show that I do not know the meaning of my vague expressions?

One way to answer depends on an externalist idea. I am party to a conceptual and linguistic practice; there are some possible differences in the practice which simply entail that my concept or language is different, regardless of how similar things may seem to me in the different situations. We cannot say in any detail what such differences are like (if we could, we would understand in detail how meaning supervenes on use, which we do not); but they are easily possible in that their obtaining would involve only minute behavioural shifts in the community, not ones with any significant impact on belief-forming mechanisms. The easy possibility is thus a social shift which drags my concept with it. What could not so easily happen is that my concept would get out of line

8 Williamson comments that he was concerned that vagueness should be a distinctive source of ignorance; thus distinct, for example, from inescapable ignorance due to limitations on accuracy of measurement. Cf. Williamson (1994b).

with the community's. The difficulty is not causal, but issues from the externalist view: what counts as the precise extension of a subject's vague concepts is fixed on a community wide basis, so, in close worlds, idiosyncrasies in his usage will not count as manifestation of an idiosyncratic concept. In close worlds, a subject's concepts coincide with those of his fellows, whether theirs are the same as or different from their (and his) actual ones.

If my concept cannot easily get out of line, then I cannot easily come to believe that "red" means *red\** when this is false. The only easy way to have the concept *red\** is to have it because the others do; but if that's the concept we all have, "red" does mean *red\**. Running the point in the other direction: if, as could easily have been, "red" meant *red\**, then my actual judging that "red" means *red* would have been a judging that "red" means *red\**, and so would have been true.<sup>9</sup>

Externalism is not the only way to attain the structure of this answer (though it seems to be the way Williamson would prefer). Perhaps I have a mechanism which reliably aligns my concepts with the ambient ones. Its reliability ensures that neither of the two possibilities which seemed to threaten knowledge of meaning would be easy; so the threat would peter out. The truth, I presume, requires a combination of both kinds of factor. There will be some externalist, thus constitutive, determinations; but these will be possible only if some causal mechanism reliably keeps me in touch with the concepts, language and topics of discourse of my fellows, for this mechanism will be relevant to *which* other speakers and objects help constitutively to determine my concepts and meanings.

## 8 Objections

1) I know that I do not have arthritis in my tibia, but an objection to the proposed account is that it entails that I do not. The inflammation I am suffering is similar in kind to that associated with arthritis, but the tibia is not a joint. However, I could easily have been wrong because it could easily have been that my community had no use for the concept of arthritis, and used instead the concept *twarthritis* which applies to any arthritis-like inflammation. If this had been so, my belief-forming episode would have issued in the falsehood that I do not have *twarthritis* in my tibia. But I know that I do not have arthritis in my tibia, so the account of knowledge is incorrect.

Being conceptually normal, I actually share my fellows' concepts and would do so if they had been different: I am actually an *arthritis* user and potentially a *twarthritis* user. In the latter case, it cannot be assumed that the (counterpart of my actual) relevant belief-forming episode would have delivered the false belief that I do not have *twarthritis* in my thigh. Actually, my

9 This retracts a suspicion voiced by Sainsbury (1995: 595–97).

belief that I do not have arthritis in my thigh comes from my doctor's explanation that arthritis does not affect joints. There can be no such explanation in the twarthritis world, since, in that world, there is no concept *arthritis* to enter into such an explanation. If my twarthritis belief comes from my doctor, it is more likely to be the true belief that I have it in my thigh than the false belief that I do not. Likewise for other aetiologies of my actual belief.

2) I know that there is a glass of water in front of me, but an objection to the envisaged strategy is that it entails that I do not. I could easily have been on twin earth, in which case it would have been twater rather than water in the glass and had I formed my belief in the way I actually did it would have been false. So although I believe truly that there is a glass of water in front of me, I could easily have been wrong, and so I do not know.

To firm up the example, let us suppose that twin earth and earth are mutually inaccessible, so that there is no question of an inhabitant of one finding himself on, or travelling to, the other. This does not settle whether I could or could not easily have been on twin earth, for that question, as the notion has been developed here, is not about ease of access but about the extent of hypothetical changes in my behaviour.

Suppose it is allowed that I could easily have been on twin earth, that is, that my belief-forming mechanism would count as operating in essentially the same way on twinearth as on earth: in both cases, the same kinds of inputs yield the same kinds of outputs. This involves thinking of inputs as the same, and so not thinking of earth inputs as responses to water and twin earth inputs as responses to twater. Perhaps the inputs should be thought of as water-appearances, to which I would be subject on either planet. If the output is thought of as a belief whose content subsumes both water and twater, then I would not have formed a false belief on twin earth, so this is not an example of an easy possibility of error. The combination of views required for the counterexample sees the outputs as the same through both being judgements that there is a glass of water in front of me.

As far as I can see, the counterexample, once it has been allowed to develop to this point, can be blocked only by denying that the twin output would have this same content; in other words, only by some form of semantic externalism. The externalist will say that the supposition that my belief-forming episode occurred on twin earth amounts (given inaccessibility) to the supposition that I was always there, and so had the concept *twater* where I actually have the concept *water*; so that episode (or its twin counterpart) would have issued in the true belief that there is a glass of twater before me.

If this is right, then something said in §2 needs to be reconsidered. As the notion of easy possibility has been developed here, my being a brain in a vat cannot be ruled not to be easily possible on the grounds that it would require difficult technology to arrange. Ease is determined by how great the behavioural changes would be. Externalistically considered, there would indeed be great differences between the vat world and the actual world: depending on

precise details of the set-up, I would likely have access to few if any of the contents I actually have, and little or none of my behaviour could be regarded as seeing or responding to barns, or whatever. But if this is the ground for saying that a vat world is not easily possible, and so is no threat to knowledge as constrained by the Reliability Conditional, then the vat world does not pose the envisaged problem for Nozickian tracking, since it verifies the counterfactual that if I were a brain in a vat I would not believe I was not one. This is because, if I were a brain in a vat, I would not, given externalism, have the conceptual resources required in order to believe anything about brains or vats.

The upshot is that if the easiness of a possibility has been rightly characterized here, then, given standard views about what is known, the Reliability Conditional leads to some form of externalism. This destination is not, of course, one that would be unwelcome to Williamson.<sup>10</sup>

10 Thanks to Timothy Williamson for commenting on the review mentioned in footnote 9, and to Keith Hossack, Christopher Hughes and Ernest Sosa for helpful discussion of the present paper. [Added note] The original printing should also have contained an acknowledgement of the influence of Peacocke's "The modality of freedom" (1998). I had heard Peacocke's lecture before I wrote this paper, but only on seeing it in print, after this paper was published, did I appreciate the extent to which it had influenced me.

# Fregean sense

---

## I Introduction

The main aim of this paper is to show that a more or less Fregean conception of sense is implicitly involved in the supposition that people are sometimes fully justified in engaging in deductive reasoning or in affirming truths of logic. Sense, as envisaged here, attaches to tokens of expressions, and sameness of sense arises when there is an immediate guarantee of sameness of reference.<sup>1</sup>

In this section I set out three background assumptions. In the following section (§2) I consider some attacks that have been made upon a Fregean conception of sense by Ruth Millikan. In the last section (§3) I apply what I hope I will by then have established to argue for my main conclusion.

### *1.1 Sense is a property of tokens*

I shall assume that the sense–reference distinction applies at the level of individual utterances: to expression tokens rather than expression types. One reason for this assumption is that it is obvious that a constant sense cannot by itself determine the reference of expression types whose tokens vary in their reference from occasion to occasion; and the determination of reference by sense is absolutely essential to any view deserving to be called Fregean.

Taking utterances as the fundamental bearers of the semantic properties requires adjusting the account of the link between sense and understanding. The basic notion is that of understanding an utterance, which can be equated with grasping its sense. Understanding an expression type is then a matter of being able to understand arbitrary utterances of it, given the required contextual information.

Taking utterances as basic makes room for the possibility that tokens of intuitively non-synonymous types should have the same sense. If on Monday I utter “It is sunny today” you can on Tuesday report me in the words “He

1 The main ideas for the paper can be traced to Campbell 1987 and Strawson 1957.



said that it was sunny yesterday”. In Fregean spirit, I assume that the relevant part of an accurate report of an utterance has the same sense as the utterance reported. So my token of “It is sunny today” has the same sense as your token of “It was sunny yesterday”, even though they are of non-synonymous types.<sup>2</sup>

Moreover, the focus on tokens makes vivid the possibility, indeed actuality, that equiform tokens differ in reference, and so highlights the theoretical need for a way of describing the cases in which equiform tokens are alike in reference, and are properly unthinkingly treated as alike. This fact is crucial for §3 below.

### **1.2 Sense without senses**

I make use of the notion of sameness of sense, but not of the notion of senses or thoughts as abstract entities. Although there may be theoretical pressures towards abstraction (for example, senses might have a special role to play in an account of indirect discourse), it seems to me useful to start simply with sameness of sense. This helps curb excessive explanatory and reductive pretensions. For example, it makes no room for the idea that senses are entities which are distinguishable from other kinds of things by the fact that we are specially good at recognizing them, at telling when we have been confronted with two different senses as opposed to the same one twice. It also discourages the idea that a sense is some psychological state which a creature could be in independently of its capacities for language or thought, a state whose nature could contribute to an explanation of why we think or understand as we do. By contrast, I shall refrain from assuming that there is any non-semantic way of specifying the relevant features of sense; perhaps there is no other fundamental and reliable way of referring to the sense of “Aristotle” than as I have just done.

### **1.3 Sense and explanation**

Can one explain how it could be that the ancient astronomers believed that Hesperus is Hesperus without believing that Hesperus is Phosphorus by alluding to the fact that “Hesperus” and “Phosphorus” do not stand in the same-sense relation? I think not. One can, indeed, infer from the (presumed) joint truth of the belief-ascriptions that there is a difference in sense in the English words, but this is a long way from connecting with the cognitive mechanisms of the early astronomers. Presumed data such as the joint truth of these belief-ascribing sentences is marked at the level of theory by ascribing distinct-

2 As Keith Hossack pointed out, Frege’s own view, in the case of indexicals, is arguably that the bearer of sense and reference is a composite of an expression type and a feature of the context, e.g. a time. (Cf. also Künné 1992 and Harcourt 1993.) I do not think that this interpretation would call for any revision of the main argument of this paper.

ness of sense. But that mark does not provide an explanation of the phenomenon except in so far as it connects this case with associated phenomena: for example, with apriori knowledge, with understanding, with rationality. In general, the explanatory value of sense lies in its putative subsumption of a number of apparently distinct phenomena. But if we take any one phenomenon, for example the one about what the ancients believed, to invoke difference of sense is simply to redescribe the phenomenon in the theorist's preferred vocabulary. It is not to acquire access to some potentially explanatory but distinct fact.

Similarly, though the main claim of this paper is that fully justified logical reasoning implies instances of the same sense relation, this is not to say that the justification of the reasoning can be explained by the relation. There will be something explanatory about the description of the situation in terms of sameness of sense if, and only if, this relation has work to do in other areas (opacity, understanding, etc.). Establishing that there is a single relation which subsumes the various phenomena, and so justifying the modest explanatory role of sense, is a task well beyond the scope of this paper.

## 2 Millikan's objections to Fregean sense

### 2.1 *Meaning rationalism*

In a series of papers (Millikan 1991, 1993a, 1993c), Millikan has adopted a position strongly opposed to Fregean sense. I find her work highly rewarding, even though my use of it consists not in accepting her criticisms, but rather in seeing how a Fregean can properly respond to them.

Her opening shot is the suggestive idea that sense is like sense data. Rather as, in perception, similarities between veridical and non-veridical cases have led people to postulate a genuinely common element, a sense datum, which is the immediate object of perception, so dissimilarities between thoughts of the same have led people to postulate genuinely distinct elements, senses, which are the immediate objects of thought. This analogy reaches its full development in her characterization of a doctrine she calls "meaning rationalism". Unlike the world of real objects, the world of meanings or senses is open to infallible knowledge by any rational mind.

The charge is laid in more concrete detail in her paper "White queen psychology" (1993c), where she characterizes the version of meaning rationalism most relevant to a Fregean by the following three doctrines:

1. **Givenness of identity and difference in mode** An intact person has the capacity to discern apriori whether two of her thoughts exemplify the same or different semantic modes of content presentation.
2. **Givenness of univocity in mode** An intact person has the capacity to discern apriori whether or not a semantic mode of content presentation for her thought is ambiguous (if ambiguity in thought is possible at all).

3. **Givenness of meaningfulness** An intact person has the capacity to discern a priori whether she is entertaining a thought or instead a meaningless form having no semantics, a mental impression that exemplifies no semantic mode of kontent presentation at all. (1993c: 326)

Here “kontent” means content in the sense of Kaplan (1977), and corresponds approximately to the Fregean notion of *Bedeutung*; so “semantic mode of kontent presentation” corresponds to Fregean sense. Her use of the word “thoughts” is not Fregean (as she notes) for two reasons: since she does not allow for a realm of sense, and thoughts are not kontents, they can only be vehicles of kontents: expressions, or their mental analogues; and she allows subsentential thoughts, counting among thoughts such things as an idea of Alice. Her use of “intact” rather than “rational” corresponds to a disambiguation she believes is required: she thinks that it would be pointless so to define “rational” that the truth of the theses is guaranteed (and then what is controversial is whether there are any rational people), and uses “intact” to stress that she is concerned with anyone whose cognitive functioning is normal without prejudice to the question whether normality invests the thinker with such powers as that of being able to identify arbitrary contradictions.

A striking feature of meaning rationalism, as embodied in these theses, is that it is a doctrine about a subject’s knowledge of semantic features of thought or language. I have not found any basis for ascribing such a position to Frege.<sup>3</sup> For example, in the opening pages of “On Sense and Reference” it is plain that the knowledge which interests him, and which will provide a way of introducing the distinction between sense and reference, is not knowledge of language or thought but knowledge of (non-linguistic) objects. In particular, arguing that identity is a relation between objects rather than between signs is tantamount to arguing that knowledge of identity, whether of the “ $a=a$ ” form or the “ $a=b$ ” form, is knowledge of objects as opposed to seman-

- 3 There are certainly passages which appear to be inconsistent with the supposition that Frege believed in infallible knowledge of sense. For example, discussing whether one could count as a definition an attempt to explain an existing simple sign by identifying its sense with that of a complex he writes:

Let us assume that A is the long-established sign (expression) whose sense we have attempted to analyse logically by constructing a complex expression that gives the analysis. Since we are not certain whether the analysis is successful, we are not prepared to present the complex expression as one that can be replaced by the simple sign A. (Frege 1914: 210)

The concluding sentence of this passage provides another example:

The senses present us with something external and because of this it is easier to comprehend how mistakes can occur than it is in the case of the logical source of knowledge which is wholly inside us and thus appears to be more proof against contamination. But appearances are deceptive. (Frege 1924/5: 269)

tic knowledge. So my first point is that whatever the defects of meaning rationalism, we have no reason to think that they bear on Fregean sense.

However, there are many arguments in Millikan which a Fregean must address. She questions whether any identity sentences can be known apriori, whether any identity sentences are genuinely informative, and whether it is ever so much as possible to invoke semantic differences to explain ignorance of an identity. In this subsection, I want to consider an argument, which I think I detect in her work, using externalism<sup>4</sup> to discredit the possibility of apriori knowledge, and hence the possibility of making the distinction between different kinds of identity sentence which Frege provides at the start of "On Sense and Reference". The use to which externalism is put blurs the distinction of the previous paragraph between the claims about semantic knowledge made by the meaning rationalist and Frege's claims about knowledge of various other matters (identity, heavenly bodies).

Given semantic externalism, whether or not an expression has meaning depends upon whether it has various causal/historical/social properties. It is not knowable apriori whether any expression has such properties. Hence it is not knowable apriori that any sentence is meaningful. Since a sentence can be true only if it is meaningful, it cannot be known apriori whether any sentence is true. Applied to Frege's distinction, we can condense the argument into the following form: given that it is not apriori that the sentence " $a=a$ " has sense, it could not be apriori that  $a=a$ , or that the sentence " $a=a$ " is true.

The objection brings out the care we must take in the present discussion over use/mention distinctions (a care which is salient in Frege's own writings). I will stipulate that a *sentence* is knowable apriori iff the knowledge involved in understanding it forms a sufficient basis on which to work out that it is true; and that it is knowable apriori *that p* iff the knowledge involved in thinking that *p* forms a sufficient basis on which to come to know that *p*. The definitions are connected by the following: if sentence *s* means that *p* then: *s* is knowable apriori iff it is knowable apriori that *p*. I will refer to the first as "sentence apriority" and to the latter as "fact apriority".

If we rephrase the objection in terms of sentence apriority, it runs: given that the sentence " $a=a$  has sense" is not apriori, how can the sentence " $a=a$ " be apriori? The answer is obvious: in the first sentence (" $a=a$  has sense"), " $a=a$ " is mentioned, not used, so understanding it does not involve understanding " $a=a$ ". In the second sentence, " $a=a$ " is used, not mentioned, so anything involved in understanding it is available as input to reasoning whose output may count as knowable apriori. Understanding " $a=a$ " is possible only if " $a$ "

4 Since Frege himself counts as a semantic externalist (if all this means is that meanings are not in the head), it might seem that a careful formulation of the relevant thesis is needed here. However, the only thing that matters to Millikan is that semantic properties should not be accessible apriori. As I will show in detail, this is something Frege has no need to deny.

has sense. If externalism is true, “*a*” has sense only if it has various causal/historical/social properties. The condition for sentence apriority, namely that it be understood, ensures that every expression used in the sentence in question has sense; so, given externalism, ensures that every expression used in the sentence and which has externalist semantics has the appropriate causal/historical/social properties. So there is nothing inconsistent in combining the claim that the sentence “‘*a=a*’ has sense” is not knowable apriori with the claim that the sentence “*a=a*” is knowable apriori. An analogous point can be made if we keep to fact apriority throughout.

In general, externalism does not threaten apriori knowledge. Rather the opposite: since understanding a sentence, or being able to think that *p*, are givens for apriori knowledge, the richer the account of understanding or thinking the greater the input, and so the greater the potential output. This shows that apriori knowledge is not to be identified with environment-independent knowledge. Apriori knowledge can be extracted from what is necessarily involved in understanding. If this is environment-dependent, then what is extracted may likewise be environment-dependent.

The argument just considered sought to make the possibility of apriori knowledge in general rest upon the possibility of apriori semantic knowledge. Once this is seen as groundless, we can reinstate the previous distinction. Whereas Millikan’s meaning rationalist is concerned with specifically semantic knowledge, the knowledge with which Frege was concerned in introducing the sense-reference distinction was certainly not semantic. We must distinguish between the following kinds of claim:

*Apriori knowledge of semantic fact:*

If *a* and *b* are (singular) expressions with the same sense and one knows the sense of each, one can know apriori that they have the same sense.

*Apriori knowledge of non-semantic fact:*

If *a* and *b* are singular expressions with the same sense, one can know apriori that *p*, where  $\hat{a} \hat{=} \hat{b}$  says that *p*.

A Fregean is certainly committed to apriori knowledge of identities, non-semantic facts. But I see no reason for him to be committed to apriori knowledge of semantic facts. Believing in Fregean sense does not involve being a meaning rationalist.<sup>5</sup>

5 In a helpful comment received after I had submitted this paper for publication, Ruth Millikan said that this section misrepresents her. She is not concerned (1993c) with knowledge of language, but with mental representations. It might be that the view here presented, in which there is Fregean sense without senses (as entities), fails to engage with Millikan’s real target. Her comment opens up avenues which I would like to explore; though this must await another occasion.

## 2.2 Identity and folders

Millikan says that:

[Frege's argument] assumes that if some sentence "*A is B*" is to effect a change in one's cognitive processes, this could only occur because it effects a change in one's bank of information (Millikan 1993c, p. 335).

She is surely right to attribute to Frege the presumption that sentences "*a=a*" and "*a=b*" can express different items of knowledge even when there is no difference at the level of reference, a presumption for which he felt no need to argue. She challenges this presumption in two ways: by offering a folder-merging account of identity, designed to make one think that identity sentences do not contain information; and by suggesting that the identity phenomena can be accommodated without sense. The first of these strategies, to be considered in this subsection, would undermine Frege's claim that a sentence of the form "*a=b*" can be informative (of considerable cognitive value), and so would undermine Frege's contrast with sentences of the form "*a=a*".

One might see value in the folder-merging metaphor without abandoning a Fregean conception of sense, for it would be consistent, given only what has so far been said, to combine the views. To prevent combination would require some principles relating folder operations to the presence or absence of information in sentences. Suppose we represent the simplest contents of folders using open atoms and their negations, counting different sentences as corresponding to distinct pieces of "folder-information". Perhaps the *a*-folder contains *Fx* and *Gx* and the *b*-folder contains *Fx* and not-*Gx* and *Hx*. Coming to accept that *a=b* is then represented by the formation of a new *alb*-folder. All pieces of folder-information from both the *a*-folder and the *b*-folder are moved to the new one (which in the example will accordingly contain all of *Fx*, *Gx*, not-*Gx* and *Hx*), and the now empty old folders are removed. Without directly contravening any Fregean view, we can suggest that coming to accept an identity as such never involves adding or deleting any piece of folder-information across the system. (Of course, in cases like the one imagined the system requires adjustment to restore consistency, but this could be seen as coming after the acceptance of the identity.) Finally, we could suggest that a sentence expresses information only if coming to accept it essentially involves adding or deleting some piece of folder-information. This rules all identities uninformative, while leaving standard examples of informative sentences as informative.

The existence of the model shows at most that we *could* take this line about the informativeness of identities.<sup>6</sup> We *must* take this line only if we must not

6 Even this may overstate the case, for perhaps coming to accept that all *Fs* are *Gs* is a matter of acquiring the disposition to add *Gx* to any folder containing *Fx*, which, if no folders contain *Fx*, does not involve any change in the pieces of folder-information across the system.

only accept the model, but must also accept its use to define informativeness on the basis of the notion of folder-information. For the first requirement, we would need a reason for not reckoning as changes in folder information those which result from restoring consistency within the new folder (the deletion of one of  $Gx$  or not- $Gx$  in the example above). For the second necessity, we would need a reason for not reckoning as informative any sentence the new acceptance of which involves some change in the system of folders. This might satisfactorily rule tautologies uninformative (supposing for a moment that these are held to be unproblematic examples of the uninformative), but it would rule identities informative. Millikan's case against Frege on the basis of the folder model has not even got off the ground.

This does not show that the case could not be made out, but optimism is diminished by the reflection that the identity phenomena are reproduced rather than annihilated in the folder metaphor: since every folder is merged with itself, there can be no question of *coming* to accept that  $a=a$ , whereas there can be such a question with respect to  $a=b$ .

### **2.3 Identity, recognition and the apriori**

Millikan challenges Frege's views about identity on another front, claiming that our inability to see that  $a=b$  should be grouped with our less than perfect capacities for recognition, and needs no explanation in terms of anything distinctively semantic. Perhaps, she suggests, the difference between " $a=a$ " and " $a=b$ " is "merely notational". Even if they are entirely alike in point of meaning, we might in the latter case fail to recognize the same meaning again. But this need no more raise a semantic issue than our inability on occasion to recognize our acquaintances.

While I agree that it is right to press this line against Frege, the anti-Fregean needs to have a positive doctrine about the possibility of apriori knowledge of identity. In Millikan's terms, the question is how the "notation" involved in " $a=a$ " can guarantee access to a piece of knowledge, whereas the notation involved in " $a=b$ " may obstruct access to what is supposedly the very same knowledge.

I shall take it for granted that Frege is right to think that some identities are knowable apriori and that some are not.<sup>7</sup> It seems to me that if one is to attempt to explain this contrast without invoking the notion of sense, one will most likely appeal to the notion of logical form, claiming that  $a=a$  is knowable apriori in virtue of its logical form, whereas  $a=b$ , is not. In §3 below, I argue that an appropriate notion of logical form requires Fregean sense. But first I will address one more anti-Fregean consideration adduced by Millikan.

7 It may be that Millikan thinks that externalism defeats the very idea of apriori knowledge, though I showed in §2.1 above that there was no easy way to establish this.

## 2.4 Psychology and semantics

Millikan rightly points out that different psychological modes of presentation cannot be assumed to be different semantic modes of presentation. I am on her side in this discussion, but, again, we draw different morals. Whereas her aim is to throw doubt on the very idea of a semantic mode of presentation, I see in such considerations reasons for not using the notion of mode of presentation as anything more than a quick initial guide to Fregean sense. As her discussion makes plain, the very phrase suggests the possibility of a kind of reduction that I think is unavailable. It suggests that a sense is contingently a sense: it is a kind of psychological state which, in general, could exist whether or not it happened to be linked with some expression in such a way as to constitute its sense. Once this approach is allowed to dominate, it becomes hard to find modes of presentation which have the intersubjective nature required for Fregean senses.

My preferred approach, by contrast, is to see senses, or more exactly the same-sense relation, as essentially linked to the description of language, with no assumption that there is a reduction to entities or relations which have a place in the description of non-linguistic phenomena. If asked to distinguish the sense of “Phosphorus” from that of “Hesperus” I would rely upon the kind of answer given by John McDowell (1977): that the senses are different is shown by the fact that one expression may be used in a correct specification of what someone said where it would be incorrect to use the other. If asked to specify the sense of “Phosphorus” I have no better simple answer than: the sense of “Phosphorus”.

## 3 Logic and guaranteed sameness of reference

“If  $p$  then  $p$ ” is supposedly a “valid logical form”, from which it should follow that every instance is true, indeed, is valid, is a truth of logic. However, the alleged instance “if John is sick then John is sick” may not be true if “sick” is understood in one of its meanings on its first occurrence, another on the second.<sup>8</sup> Logical forms presuppose a pattern of recurrence: they assume that for each schematic letter, each token of it is to be understood as making the same contribution to logical truth and validity. If we are to recognize a genuine instance of a logical form as valid, we must be able to recognize that replacements of equiform tokens of schematic letters make equal contributions. Hence the concept of logical validity requires the notion of a relation

8 Cf. Strawson, 1957; and, for an application of the point close to that made here, Campbell 1987. “Sick” is not shown to be ambiguous by the fact that one can be sick in body without being sick in mind, or vice versa, any more than “human” is shown to be ambiguous by the fact that a human can be a man without being a woman, or vice versa. But Strawson’s point is good, even if the example is a little infelicitous.



between tokens which obtains only on condition that they make the same contribution to validity, and whose obtaining can be manifest to reasoners. In this section, I argue that one constraint upon sameness of sense among singular tokens is that it should make sameness of reference, and thus sameness of contribution to validity, recognizable.

The phenomenon at issue has not received much attention, in part because logic operates under the assumption that expression types are the prime bearers of logically relevant properties, and that for tokens to be of the same type is a merely spatial issue, raising no epistemic problems. Even if the last claim were true, which is doubtful, this paper runs on the assumption that sense is in the first instance a property of tokens. This makes room for the observation that if one counts anything an instance of " $a=a$ " if it has equiform tokens either side of the identity sign, not every instance is true. Moreover, even if the tokens have the same reference, this is not enough for the identity sentence to be knowable a priori, as "Paderewski" examples demonstrate.<sup>9</sup> We have a priori knowledge in such cases only if something in the understanding of the sentence guarantees its truth. Hence a priori knowledge requires that something in the understanding of the sentence guarantees that the tokens have the same reference. I use "sense" to mark this relation between tokens. The conclusion is that logical notions such as "logical form", "logical truth", "logical validity", cannot be understood in the way they are intended without presupposing a notion of sense. Hence one does not avoid sense by claiming that the distinction between the a priori knowability of  $a=a$  and the non a priori knowability of  $a=b$  is based on their distinct logical forms.

It is a premise to this argument that we are sometimes completely justified in affirming a truth of logic or using a logical pattern of argument. I claim that this is possible only if we are, in those cases, completely justified in unthinkingly taking it for granted that recurring tokens of a given type make the same contribution to truth (in the case of singular terms, have the same reference). When this happens, I say that the similar tokens have the same sense. This does not explain how we are justified, but merely marks the fact that we are. There may be no general explanation of how we are justified in such cases.<sup>10</sup>

To illustrate this conclusion, consider an argument (that is a series of genuine – interpreted – sentences) of the following form:

9 Suppose a person thinks that there's Paderewski the statesman and also a different Paderewski, the pianist. He is wrong because in fact the statesman is the pianist: there is just one Paderewski. Yet such a person can, without obvious irrationality, believe something he could express as "Paderewski is not Paderewski". Cf. Kripke 1979.

10 For example, the basis for the justification of reasoning involving repeated perceptually based demonstrations of an object (cf. Campbell 1987) may be different from that involving propositionally valid reasoning.

$$\begin{array}{l} Fa \\ a=b \\ \hline Fb. \end{array}$$

On a standard reading, we just assume that tokens of equiform expressions make the same contribution to validity. However, when we consider that pairs of equiform tokens may differ in meaning, in a way that would undermine the validity of an otherwise valid argument, we appreciate that this assumption does not do justice to the realities. Let us mark distinct tokens of the same type with a different number of dashes. Then a more illuminating presentation of the argument form would be something like:

$$\begin{array}{l} F'a' \\ a''=b' \\ \hline F''b''. \end{array}$$

This way of putting it makes it plain that there is a real question about whether the occurrence of “*a*” on the first line has the same reference as the occurrence of “*a*’” on the second; and likewise in the other cases. My claim is that sometimes this identity is guaranteed by what is involved in understanding; in these cases, we say that the relevant tokens have the same sense.

Some identities are never fully expressible without semantic ascent. Suppose that, in the above example, we try to express what, in the previous paragraph, I expressed metalinguistically, by writing:  $a'=a''$ . This is strictly nonsense, for these marks are supposed to be tokens, and the first token of “*a*’”, marked by appending just one prime, occurred in the first line of the argument whose form was displayed a few lines up, and it can never recur anywhere else. I can refer to that token, and say that it has the same reference as the second token of “*a*’”, the one on the line below the first, but it is impossible to say this by re-using these tokens: they have been used up. This means that we cannot regard arguments like the one under consideration as enthymatic, needing but a further (object-language) sentence to be made completely valid; there is no evading unthinking reliance on sameness of reference. It also gives us a deeper understanding of the low cognitive value of a sentence of the form “ $a=a$ ”, in our terminology, “ $a''=a'''$ ”. Every identity sentence has distinct tokens, so unless identity of reference has already been made available, no identity can be “true in virtue of form”.

The phenomenon is entirely general. Consider, for example, modus ponens reasoning, of the form “*p*; if *p* then *q*; so *q*”. If we are ever justified in using such reasoning, we must be justified in supposing that the various occurrences of *p* and *q* make the same contribution to validity. We are sometimes thus justified. The Fregean theorist I envisage will mark such cases by counting the relevant tokens of *p* as alike in sense; likewise for the tokens of *q*.

As I see it, the theorist begins by finding cases in which reasoning is indisputably both justified and correct. Acknowledging this justified correctness at the theoretical level involves seeing reasoners as justified without further ado, and also correct, in treating tokens as having the same reference. The theorist is to record this fact by counting the tokens as alike in sense. In the first instance, there is no need to think of this record as explanatory. It gains explanatory status if the same-sense relation has other work to do in the theory (as I believe it has): in describing indirect discourse, for example. Then the explanation is not so much: “the same-sense relation obtained between these tokens; so the reasoner was justified and correct”, but rather “the reasoner was justified and correct, so the same-sense relation obtains, a relation whose role and working connect this case of correct reasoning with other phenomena (like indirect discourse)”.

Sense is not the only possible guarantor of sameness of reference. Anaphor also plays this role. If we try to use this notion to give a representation of our sample inference, we might write:

$$\begin{array}{l} a \text{ is } F \\ \text{it}_a = b \\ \hline \text{it}_b \text{ has that property}_F. \end{array}$$

Here the subscripts serve to make explicit the anaphoric heads of the expressions to which they are attached. However, in inference we need transportable conclusions, ones which can be placed in other contexts from which the relevant head may be absent. This involves our using further inferential steps, for example, turning the above conclusion into “*b* is *F*”, and these further steps raise the same issues as before, requiring that it be legitimate for a speaker to take it, without evidential support, but merely on the basis of his understanding of the expressions, that two tokens have the same reference. So we cannot use anaphora to dispense with recognizing the phenomenon of guaranteed sameness of reference; that is, of sameness of sense.<sup>11</sup>

11 Versions of this paper were read at Logica 1996 in the Czech Republic, at Sheffield University’s Philosophy Department Seminar, at a one-day conference on Frege organized by the Philosophy Programme of London University’s School of Advanced Study and at the Philosophy Society of Portugal in Lisbon. I would like to thank participants for the many helpful comments and suggestions I received on those occasions. I would also like to thank the following for detailed written comments which I found extremely valuable: Michael Beaney (whose book, cited above, appeared too late for me to be able to take it into account), Keith Hossack (my commentator at the London conference just mentioned), Michael Martin, Christopher Peacocke, and Tom Stoneham. After I had submitted this paper for publication, I had occasion to show it to Ruth Millikan, who thought that I had misunderstood her on various points. I have briefly addressed the claimed misunderstanding she thought most serious in footnote 5 above, but printers’ deadlines made it impossible to do justice to her comments.

# Indexicals and reported speech

---

## I A hypothesis about meaning

If I understand an utterance whereby someone says something, I know what was said, and typically I can express this knowledge: I can report what was said. If meaning is the least that must necessarily be accessed in understanding, then meaning is specified when speech is reported. The aim of this paper is not to argue for this natural hypothesis, but to explore its consequences: the consequences of treating constraints on reporting speech as guides to meaning.

Despite its vagueness, the consequences of the hypothesis are quite rich. To give a preliminary indication: if you, addressing me, utter the words “You are a fool” I can correctly report you as having said that I am a fool. The hypothesis under discussion is to be interpreted so as to entail that, since your token of “you” is properly reported by my token of “I”, these tokens have the same meaning. In general, the hypothesis entails that non-synonymous types of expressions may have synonymous tokens. This consequence may occasion somewhat more surprise than the hypothesis from which it flows. The underlying phenomenon is very familiar; the slight surprise is explained by the fact that in the dominant tradition, meaning is linked to expression types rather than to tokens. Tokens like the pair in the example, which the present hypothesis requires us to count as synonymous, are more usually regarded as related merely by sameness of reference. Yet this relation may hold between tokens one of which cannot be used to report a use of the other. For example, if the words you address to me are “You are a fool”, I cannot adequately report you as having said that the Editor of *Mind* is a fool.

Further consequences to be drawn in this paper relate to formal semantic theories (§2), and to various claims about the special role of indexicals with respect to action, to science, to time, and to the coherence of an omniscient but eternal God (§4). On the way, I find it necessary to raise some general issues about the relationship between indexical and anaphoric uses of pronouns (§3). In the final section of the paper, I raise the question of how the perspective offered relates to Frege’s views (§5). In the remainder of the

present section, I say a little more about the notion of a fully adequate and explicit report of speech, and thus make the hypothesis to be explored somewhat more precise.

In a simplified case in which we bracket context-dependence, ambiguity and actual or possible differences of language, a reporter could do no better than repeat the original speaker's words after uttering "So-and-so said that". Homophony rules. In this situation, our hypothesis, though not an incorrect guide, is uninteresting. It rises little beyond the claim that words mean what they mean. The guide becomes interesting only to the extent that we relax the simplifications. In this paper, the feature to be relaxed, at least in part, is context-dependence. I shall explore the results the guide delivers about reports of, and hence meanings of, indexicals.

The guiding thought is not that we can use intuitions about attributions of any propositional attitudes as guides to meaning. Reported speech has a special claim to be a good guide, since what is reported is essentially a language-related act, with other pressures on the utility of the report, like explanatory adequacy within some belief-desire schema, reduced to a minimum. Some difficulties about deciding whether speech has been correctly reported can be allowed as manifestations of some indeterminacy in meaning. There are also ways in which reports may be less than adequate which I wish to resolve by stipulation.

Someone utters the words "It rained every day on my holiday". Perhaps we should not count "what he said" as false if there was one day which, though misty, strictly speaking lacked rain. Taking what he said strictly, however, it was false. Someone utters the words "Travelling to Italy will cost you an arm and a leg". He has said, figuratively, that travelling to Italy will cost you a lot. But speaking literally he has said that travelling to Italy will cost you an arm and a leg. The dean utters the words "I believed that Professor Z was overpaid". The dean has just discovered that he is Professor Z, whom, under that anonymous label, he has just placed in the category of overpaid staff. (The example is a variant of one given by Perry 1983: 110.) He has said, implicitly, that he believed that he was overpaid, but he has not said this explicitly.

There are other ways in which a report of speech may not be fully explicit. In some contexts, even a correct specification of the actual words uttered may be esteemed defective, either because the audience does not understand the relevant language, or because knowing the words uttered may be insufficient to know what was said. Another kind of example of failure of full explicitness are oblique specifications of what was said; for example: James said the same as Mary.

In what follows, I shall assume that "said" abbreviates "strictly, literally and fully explicitly said". This requirement rules out various kinds of cases which might otherwise make the hypothesis that we can use speech reports as guides to meaning so implausible as not to be worth much exploration. If someone says of you that you are a fool, I can arguably report this to you correctly by

saying “He said that you are a fool”. On the face of it, this does not imply anything about the means he used to refer to you. Yet it would be highly implausible to suppose that the occurrence of “you” in my mouth was synonymous with the words the original speaker used to refer to you, whatever these words may have been. Such a result would make it miraculous that my report could be understood.

The situation is not remedied by regarding the logical form of such reports as *de re*, for there will still be an expression which refers to the relevant object, and a subsequent pronoun which also does, for example: “Referring to you, he said that you are a fool”. The relevant use of such words is one in which the two occurrences of “you” have the same reference, whether because the second is anaphorically dependent upon the first, or because you are demonstrated in association with each occurrence. On the present approach, we would seem still to get the result that my token “you are a fool” means the same (assuming my report is accurate) as the speaker’s words, whatever words he used to refer to you. Similarly, if I overhear someone uttering the words “The first speaker at today’s conference is a fool” it seems that I can correctly report having overheard someone say that I was a fool (for I am the first speaker, and I know it). But it seems highly implausible to suppose that my token of “I” has the same meaning as that token of “the first speaker”. Once again, the supposition would make it miraculous that my report could be understood.

I regard these cases as reports which are not fully explicit. The default position for full explicitness is re-use of the words the speaker used, as in the simplified homophonic case. Reporting indexical speech forces a departure from this default: we cannot improve on the anaphoric style of report. If the original words were “He is a fool”, said of you, then my report “He said that you were a fool” does count as maximally explicit; but in this case the consequent identification of the meaning of the tokens of “he” and “you” is not implausible. In the other cases, however, those which seemed to reveal a very implausible consequence of our guiding hypothesis, there is more information to impart, which could be imparted in *oratio obliqua*. Hence the reports which seemed to render that hypothesis not worth exploring are not maximally explicit. It would not be surprising if reports of speech which were not fully explicit failed to coincide in their meaning with what they reported.

Within the simple picture of a single language and no indexicality, one who gave an adequate report could make himself a samesayer with the original speaker merely by uttering the words which follow “said that” in his report. In the next section, I argue that when we drop the simplification we find that one who can adequately report may not be able to say anything which makes him (save in the context of the report) samesay with the original speaker. I explore a consequence of this for systematic semantic theory.

## 2 Non-detachability and semantic theory

Here are some examples of utterances which can be adequately reported, but whose content the reporter cannot express by means of a self-standing utterance. On 12 May 1968, the revolutionary leader said that that day marked a new dawn. We can divide the report into two parts: first, the scene is set by the words “On 12 May 1968, the revolutionary leader said that”; then the content is ascribed by the words “that day marked a new dawn”. Intuitively, the report passes muster; we tend to imagine that the leader used some word like “aujourd’hui”. But one cannot simply detach the content-ascribing part from the scene-setting part, since the one depends anaphorically on the other. Without some special contextual background, of the kind supplied by the scene-setting, I cannot make myself a samesayer with the leader by uttering “That day marked a new dawn”. Indeed, treated as self-standing, that is, as not governed, implicitly or explicitly, by the kind of contextual material found in the scene-setting, I cannot say anything by an utterance of that sentence. Nor will it do to switch either to a context-independent expression, or to one dependent upon my actual context. If I utter “12 May 1968 marked a new dawn”, or “The day 8030 days before today marked a new dawn” I do not say what the leader said, if he used a word like “aujourd’hui”. Understanding my words requires knowing what the date was on the day of the reported utterance, or knowing how many days before my report it occurred, whereas understanding the revolutionary leader required no such thing. Hence these proposed words do not mean what the leader’s words meant, and they could not be used by me to samesay with him.

Frau Lauben told Dr Lauben to his face that he was wounded. “To his face” ensures that she treated him as addressee, using a word like “Du”. But I, in telling you this, cannot address Dr Lauben. If I utter “Dr Lauben was wounded” I do not make myself a samesayer with Frau Lauben, because I do not capture the feature of her content that corresponds to her use of the second person. Likewise, to report her as having said that Dr Lauben was wounded is not entirely accurate. We must stick to a two-part form, involving scene-setting and anaphorically dependent content ascription, undetached and undetachable; for example: Frau Lauben told Dr Lauben to his face that he was wounded.

We are walking in the woods and you fleetingly glimpse what in fact is a rabbit. I see it too, and know that it is the object to which you refer in uttering “That was a bear”. I can report the incident later as follows: “Seeing a rabbit, he said that it was a bear”. But on that later occasion, I cannot detach, in such a way as to utter something which would make us samesayers. I cannot utter “It was a bear” or “That was a bear”, as my current context would supply either no reference or the wrong one. And attempts like “The rabbit he saw was a bear” distort your thought. Either way, no detached yet adequate and accurate report seems to be possible.

In many cases, original utterance and report are linked in the following way: a demonstrative pronoun in the original becomes a suitably transformed anaphoric pronoun in the report. (This suggests that there should be both demonstrative and anaphoric uses of “I”, just as there are such pairs of uses of the other personal pronouns. The basis for this view would be that anaphora is at the level of logical form, for then one token of “I” could be antecedent in such a form and another an anaphoric dependent.) Where the anaphora is essential, the reporter cannot detach the content-ascribing part of the report from the scene-setting part, since this deprives the anaphoric pronoun of its antecedent. But the attempt to use any other kind of referring device in its stead may distort the content of the original utterance.

The guiding hypothesis – that we can use correctness of reported speech as a guide to meaning – raises many issues, some of which will be taken up later (especially in §3). The matter I wish to discuss now can bracket many of these, for all I ask is that it be granted that content-ascription cannot always be detached from scene-setting; in other words, that the reality of non-detachability be granted. The guiding hypothesis has it that in an adequate report of speech, a reporter specifies the content of the utterance. I shall assume that a theory of meaning for a language should do the same: given a suitable description of an utterance as input, it should deliver a specification of the utterance’s content. We can infer that a correct theory of meaning has the analogue of non-detachability. Without prejudice to the correct approach to theories of meaning, let us use truth theories as an example. Non-detachability entails that we cannot derive an interpretative T-sentence for every utterance. Consider, for example, the revolutionary leader’s utterance on 12 May 1968, call it *u*, whatever words he produced thanks to which we can properly report him as having said that that day marked a new dawn. A T-sentence for *u* has the form

*u* is true iff *p*

where *p* is, by the requirements of the biconditional, a self-standing sentence, one with a truth value in its own right. The previous considerations show that no such sentence is usable in an adequate report of what was said by *u*; hence, by the proposed methodology of being guided to meaning by how speech is reported, there is no correct T-sentence for *u*.

This has been widely recognised, though perhaps not for quite these reasons. In response, several authors have proposed that a truth theory should use clauses which are conditional in form: the antecedent specifies an utterance along with features of its context, and the consequent is a biconditional resembling a T-sentence save that its components may contain variables bound by material in the antecedent. I want to show that this approach faces a dilemma: either it delivers results inconsistent with non-detachability, or it makes it impossible for us to bring to bear the specific information we



have about an utterance in such a way as to have a chance of extracting an interpretation. Towards the end of the section, I offer a way out of the dilemma.

It will be useful to have a specific conditional truth-theoretic clause for discussion:

If  $u$  is an utterance of “today is July 4” by  $s$ , and  $s$  refers with the utterance of “today” therein to  $\delta$ , then  
 $u$  is true  $\leftrightarrow \delta$  is July 4. (Higginbotham 1994: 94)

This clause is still too general to supply an interpretation, without feeding it supplementary information. If we know that, for example, Gareth Evans uttered “today is July 4” on 4 July 1968, we would hope to be able to feed the generalization this specific knowledge in order to derive an interpretation of the specific remark. Using  $u_1$  to refer to Evans’s utterance, we may instantiate with respect to the variables  $u$  and  $\delta$ , which are implicitly universally quantified, deriving

If  $u_1$  is an utterance of “today is July 4” by Gareth Evans, and Gareth Evans refers with the utterance of “today” therein to July 4, then  
 $u_1$  is true  $\leftrightarrow$  July 4 is July 4.

The antecedent is true; so, whether we want it or not, the theory, together with an appropriate specification of the utterance, entails the consequent ( $u_1$  is true  $\leftrightarrow$  July 4 is July 4). So the utterance is being interpreted by the self-standing sentence “July 4 is July 4”. According to the thesis of non-detachability, interpretations of this kind cannot in general be correct.

The dilemma we face is not special to Higginbotham’s theory or to truth theories. It constrains any systematic attempt to provide in a compact (and thus theoretical) form, information sufficient for interpretation (given an appropriate specification of the utterance).

The difficulty I am raising must be distinguished from the complaint that anything which enables us to derive “ $u_1$  is true  $\leftrightarrow$  July 4 is July 4” must be incorrect, since this biconditional associates with  $u_1$  an uninterpretative truth condition. I want no truck with this complaint for two reasons: first, it is local to truth theories, whereas the point about non-detachability extends to any form of semantic theory. Secondly, it is not decisive even against truth theories. It is no better than the objection that from standard truth-theoretic clauses, together with the assumption “Hesperus is Phosphorus”, one can derive not only the interpretative T-sentence

“Hesperus is Phosphorus” is true iff Hesperus is Phosphorus

but also the uninterpretative

“Hesperus is Phosphorus” is true iff Hesperus is Hesperus.

The standard response, whose adequacy I shall not challenge, is that it should not be surprising that supplementing semantic information with non-semantic information should yield something which cannot itself be classified as semantic information (an “uninterpretative” T-sentence). Applied specifically to the case of indexicals, it is consistent to claim both that Higginbotham’s conditional T-sentence correctly specifies the meaning of utterances, and that consequences of this conditional, derived by applying non-semantic information to effect detachment, do not.

The complaint thus requires as a supplementary premise that the information about specific utterances of the kind we bring to bear in interpretation (e.g. the day on which the utterance was made) is specifically *semantic* information. The dominant tradition has it that it is not semantic information, for semantic information is conceived as the minimum a semantic theory should state. According to this tradition, semantic theory achieves generality by finding semantic types (“expression types”) whereby it can speak of all tokens of the type; and these generalizations will not involve any information as specific as that required to derive the contested T-sentence. So the envisaged complaint will, I believe, bog down in a messy and probably inconclusive discussion about what is to count as semantic information.

If I am right, non-detachability poses a problem for any semantic theory. Either the theory does not have a way in which our specific information about utterances, information required for interpretation, can be brought to bear, in which case it fails in its overall aim; or else, bringing the information to bear immediately yields something inconsistent with non-detachability. A possible way through the dilemma is to recognize a special kind of instantiation, which I will call “anaphora-preserving instantiation”. An example of its form could be written:

From “All  $F$ s are  $G$ ” infer “If  $a$  is an  $F$  then (that  $F$ ) <sub>$a$</sub>  is  $G$ ”.

Here the subscript  $a$  indicates the anaphoric dependence of the associated occurrence of “that  $F$ ” on a previous occurrence of  $a$ . Subscript  $a$  is not a referring expression. The predicate  $F$ , presumed to be sortal, has been carried forward to form part of the anaphorically dependent expression (that  $F$ ) <sub>$a$</sub> . This is not an essential feature of the proposal, and I offer no argument for it. Applying the idea to our particular case yields:

If  $u_1$  is an utterance of “today is July 4” by Gareth Evans, and Gareth Evans refers with the utterance of “today” therein to July 4, then  
 $u_1$  is true  $\leftrightarrow$  (that day)<sub>July 4</sub> is July 4.

This is not very idiomatic, but it seems to me to contain essentially the right

idea. It approximates the natural report: On 4 July 1968, Gareth Evans said that it was July 4. It faithfully follows the contours of the account of meaning that would result were one to suppose that meaning can be properly identified in terms of the correctness of reported speech. In particular, we have brought our specific knowledge to bear, while not detaching.

I turn in a moment to the question whether any independent motivation could be discovered for recognizing such a species of instantiation. A more immediate worry is whether the proposal, however motivated, is of any help. With certain qualifications, a sentence containing a pronoun in an extensional position whose antecedent is a singular term entails a corresponding sentence in which the pronoun is replaced by its antecedent. (Qualifications are needed so as not to validate, for example, the inference from “Only Satan loves himself” to “Only Satan loves Satan”.) The rule invites us to derive, from the clause just displayed,

If  $u_1$  is an utterance of “today is July 4” by Gareth Evans, and Gareth Evans refers with the utterance of “today” therein to July 4, then  
 $u_1$  is true  $\leftrightarrow$  July 4 is July 4.

Given our specific information about  $u_1$ , this entails the biconditional with the detachable right hand side, and we seem to be back where we started.

We need to distinguish features local to truth theories from more general considerations. The canvassed rule for pronoun replacement does not apply to a semantic theory whose syntax is non-extensional at the point of delivery of interpretations, that is, at the point corresponding to “is true iff”. The rule’s restriction to extensional positions cannot be simply deleted, if its correctness is to be preserved. For example, there is no sound inference from “On 12 May 1968, the revolutionary leader said that that day marked a new dawn” to “On 12 May 1968, the revolutionary leader said that 12 May 1968 marked a new dawn”. So we have not yet been given a reason to fear the unwanted inference in semantic theories in general.

In the setting of truth-theory, the unwanted inference is a manifestation of a familiar phenomenon which the notion of “canonical proof” is supposed to address. The problem we are confronting is no graver than the fact that a classical truth theory containing the interpretative theorem

“Snow is white” is true iff snow is white

cannot avoid also containing the non-interpretative theorem

“Snow is white” is true iff snow is white and either there are dragons or there are not.

A truth theorist will say that while the first theorem is susceptible of canonical

proof, the second is not. If this reply is adequate to the case of the unwanted conjunct, it is adequate to the case of the unwanted detachable theorem; if it is inadequate in the former case, then truth theories are, independently of the concerns of this paper, inadequate as semantic theories.

The question whether there is any independent justification for recognizing anaphora-preserving instantiation is more complicated. In classical formal languages, sameness of reference of tokens is guaranteed by sameness of singular term tokened, where equiform tokens count as tokens of the same term. This simple picture does not apply to natural languages: "Schnabel is a pianist" and "Schnabel is not a pianist" may both be true; hence equiform tokens of "Schnabel" are not guaranteed to have the same reference; hence either equiformity is not enough for being tokens of the same singular term, or being tokens of the same singular term is not enough for co-reference. Anaphora provides a guarantee of co-reference which is immune to this feature of natural language. We would wish every instance of "All pianists are pianists" to be a truth; yet "If Schnabel is a pianist then Schnabel is a pianist" is not a truth, if the first occurrence of the name refers to the famous pianist, the second to the famous artist (who is not also a pianist). By contrast, "If Schnabel is a pianist then he is a pianist" is not subject to these vagaries.

Recognizing anaphora-preservation in logical form thus has a point, even when the anaphoric dependence is upon a singular term, specifically, a proper name. The point may easily disperse, however, when further inferential needs are recognized. If we start with "All pianists are musical" we want to be able to derive that Schnabel is musical given that Schnabel is a pianist, and this appears to allow room for the vagaries which anaphora prevents. The problem could be put as a dilemma: either we have, for anaphors whose antecedents are singular terms, the pronoun replacement rule canvassed above, in which case, since the anaphoric occurrence entails a non-anaphoric one, the former is inessential and needs no special recognition; or else we do not have the pronoun replacement rule, in which case recognition of anaphora blocks intuitively correct reasoning.

Neither horn is decisive. Arguably, later occurrences of names may count in appropriate contexts as anaphoric upon earlier ones, so that one would still need to recognize anaphoric dependence at the level of logical form. Moreover, the needed restrictions on the pronoun replacement rule even in some extensional cases suggest that anaphora-preservation is not always otiose. As Geach has stressed (1962: 132, 138), if "himself" is anaphorically dependent upon "Satan" in "Only Satan loves himself", it is an example of indispensable anaphora, dependent upon a singular term.

It is often said that bound variables are like anaphoric pronouns. We have in effect uncovered a limitation to this analogy in the classical conception of instantiation, and proposed a way of restoring it to full strength.

### 3 Indexicality, anaphora and reducibility

We have, I believe, been able to derive some value from our guiding hypothesis, even though it has not been described in much detail. This section aims to describe it more fully.

In using an indexical, one exploits a perspective on the world. One locates an object by reference, ultimately, to one's own position in space or time. In interpreting a use of an indexical, one needs to locate its user's perspective within one's own. Because he was speaking on 12 May 1968, that is the day that would count as "today" for him; because she was addressing Dr Lauben, that is who would count as "you" for her; because it was a rabbit he saw and which prompted his remark, that was what counted as "that" for him. One needs to identify the perspective, not suppress it.

There are systematic transformations: you report a use of "you" as "I" if you are the addressee, as "he" or "she" otherwise; you report a use of "I" by using "he" or "she" (with optional "himself" or "herself"<sup>1</sup>); you report a use of "today" on the following day by using "yesterday". The last example is a case in which detachment is possible. "Yesterday" has the special role of being substitutable for an expression of the form "(that day)<sub>x</sub>" where the anaphoric dependence marked by *x* is on an expression which the reporter of the speech could use to refer to a day which, from his perspective, is yesterday; which in turn requires the presence of at least implicit reference to a day which the reporter could refer to as "today". Yesterday is yesterday only from the standpoint of today. This seems like an ad hoc convention, which does not run deep. (A more detailed account of "yesterday" might well treat the logical form of "John said that it was fine yesterday" as "Yesterday, John said that it was fine then". This would bring the report into line with ones using dates. The ad hoc convention allows replacement of the anaphoric "then" by self-standing "yesterday".) There is nothing odd about not using "The day before the day before yesterday" to report a "today" utterance made the day before the day before yesterday. Sometimes the expressions standardly used to report indexicals function genuinely as indexicals (e.g. "yesterday"); but more often, as in many uses of "he" and "it" ("Seeing a rabbit, he said that it was a bear"), they function as anaphoric pronouns. The reference of an indexical is governed by extra-linguistic material, of an anaphor by linguistic material. In reporting utterances of indexicals, indexicality often transmutes to anaphora. Non-detachability reflects the fact that

1 Castañeda (1966 and subsequent papers) has argued that some reports require the anaphoric use of "he himself" and the like in order to do justice to tokens of "I". This may well be so, though whether the phenomenon should be classified as semantic or pragmatic is harder to decide. The present paper proceeds on the assumption that it is pragmatic; but adding it as a semantic feature would not affect any of the claims for which I argue.

we sometimes have to use linguistic contexts to do for our report what the original utterer could rely upon non-linguistic material to do. If one regards anaphora as other than a species of indexicality, one will think that an indexical token can be properly reported by a non-indexical one. At least some indexicals would, on this view, be in a sense “reducible”. Indeed, in the next section (§4) I will go further, and claim that all indexicals are in this sense reducible.

This claim presupposes that one can properly distinguish between indexicality and anaphora, a presupposition to be examined later in this section. Before doing this, some further clarification of the guiding hypothesis (the hypothesis that we can use fully explicit reports of speech as guides to meaning) is called for.

The hypothesis falls between two extreme treatments of the meanings of indexicals. At one extreme, the meaning of an indexical token is identified with a complex description; at another, with the object of reference. The present proposal is obviously distinct from the first extreme, since it would be wrong to identify the meaning (as opposed to the reference) of an anaphoric pronoun with that of its antecedent. We have already seen that to replace the pronoun in the “said that” context with its antecedent would sometimes turn a truth into a falsehood. It is also distinct from the second, for there are constraints on the form that a maximally explicit report can take: it is not enough merely to refer to the object the original speaker referred to. If you say truly “That is Hesperus” it is not right to report you as having said that Hesperus is Hesperus.

Another version of the second extreme sees indexical uses of pronouns as requiring *de re* forms of report. In the example just given, it may be suggested that a proper report is: “Of Hesperus, he said that it is Hesperus”. Concerning the three examples of non-detachability, it would be claimed that the question whether they are correct is equivalent to the question whether or not their *de re* counterparts are correct: “Speaking of 12 May 1968, he said that it marked a new dawn”, etc. This does not do justice to the fact that whereas the *de dicto* report counts as fully explicit, arguably no *de re* report can be this, since such a report deliberately distances itself from any information about how the speaker referred to the object. Another way to put the contrast is like this: from the supposition that the *de dicto* report is fully explicit one can infer that the original speaker used a context-dependent indexical mode of referring. No such conclusion is warranted by a corresponding assumption about a *de re* report.

That one can use anaphoric tokens to report indexical ones does not imply the converse. That is good, since the converse appears to be false. Consider the exchange:

- A: I spent two hours talking to Max today.  
 B: I imagine he talked about Frege.

Suppose that B has no idea who Max is, so that we do best to regard his use of “he” as anaphoric on A’s use of “Max”. If Max is present on some later occasion, and I demonstrate him, it would be incorrect to report B as having said that he imagined that *that student* had talked about Frege.

In the first two examples in §2, the scene-setting includes reference to a day and to a person, whereas in the third example – “Seeing a rabbit, he said that it was a bear” – there is just an existential quantification. This might prompt the following objection: “Existential quantification is not reference, so a pronoun anaphoric on an existential quantification cannot be regarded as a genuine referring expression. The utterance reported in the rabbit example involved a demonstrative pronoun: a genuine referring expression. But it is absurd to suppose that a referring expression could have the same meaning as an expression which is not a referring expression, the ‘it’ in the report.”

One line of reply would be to accept the argument as sound, and say that the report in the rabbit example is not fully explicit, since it has not pinned down what the original speaker referred to. I prefer an alternative approach. Philosophers have been brought up to be highly sensitive to the contrast between quantification and reference; otherwise, for example, they cannot engage in discussions of Russell’s theory of descriptions. I do not wish to dispute that a distinction can be drawn which is crucial for some purposes. But it is not obvious that it produces the kind of distinction which can properly be used in the envisaged objection. Many entities are introduced into discourse under existential quantification, and we speak of “reference” without imputing to interpreters any capacity for unique identification. A historian who begins “The King had a sister who was a great comfort to him. When times were hard, she . . .” engages in an entirely banal form of speech, yet arguably, regardless of how many sisters the King had, can go on properly to use referring expressions for just one sister, which the novice audience can understand. It would be arbitrary to count the later indisputable referring expressions as such, yet refuse this reading of the first “she”. Similarly, “The King had a sister called ‘Matilda’ . . .” arguably puts a novice audience in a position to use “Matilda”; and this would normally count as a referring expression, even though it is patently introduced on the back of mere existential quantification. By these possibly lax standards, the anaphoric “it” in the rabbit example counts as a referring expression. The objection of the previous paragraph would need to establish that this token’s synonymy with a demonstrative requires it to be a referring expression according to some more demanding standard.

The discussion so far has taken for granted that there is a clear and firm distinction between indexical uses of pronouns, e.g. the use of demonstratives, and anaphoric ones. The standard claim is that indexicality draws on features of the token’s non-linguistic context, whereas anaphora draws on features of the token’s linguistic context. A further generally recognized distinction is that anaphoric dependence can guarantee co-reference, in a

way in which recurrent demonstratives cannot, even if they in fact co-refer. However, from some points of view, perhaps that of the psychological skills needed to engage in linguistic activity, there may seem to be something more like a spectrum. Cases of deferred demonstration are in some respects intermediate. Suppose that a silent movie camera was running in the Lauben residence just when Frau Lauben was telling Dr Lauben to his face that he was wounded. If, many years later, I arrange for us to watch the movie together I can say: "Now she's saying that he's wounded". Intuitively, my token of "he" is demonstrative: it refers to Dr Lauben via referring to his image on the screen. It also counts as demonstrative rather than anaphoric by the standard test, since there is no relevant linguistic context. Yet it also seems hard to deny that in some respects it resembles the anaphoric case: the role of the movie is analogous to the role of scene-setting remarks.

From some points of view, an important distinction would be between naturally occurring and deliberately contrived contextual features. The paradigm of the deliberately contrived is language, as in the verbal scene-settings considered in the previous sections; less common cases would be use of images and other traces or icons. What is interesting about the contrived cases is that we take control: we transform the situation in which we find ourselves into one related in a special way to the original speaker's situation, a way which enables us adequately to report what he said.

The non-detachability thesis could be seen as expressing merely the thought that there are cases in which I cannot accurately report without contriving contextual features for the special purpose of giving my words the right reference. This weaker thesis would be strong enough for the purposes of the previous section (§2). But a feature of language is that, if we know what was said, there are words which will effect the appropriate scene-setting, and there are cases in which (in the absence of appropriate images or whatever) only words can do this. With this in mind, a stronger thesis of non-detachability is available: there are situations in which reporters have to engage in verbal scene-setting if they are to report an utterance correctly. In these cases, anaphora is indispensable.

## **4 Reducibility: action, science, time and God**

### ***Reducibility***

In some cases, anaphora is indispensable; but in all cases, it is available as a resource in facilitating reports. This last claim is another way of expressing the thesis that all indexical tokens are reducible: indexical speech can always be reported by words which, in their content-ascribing role, are not indexical (though they may be anaphoric).

The thesis does not extend to the scene-setting component of a report of speech. For all I know, in order to set scenes we may, sometimes or always,



need indexicals either explicitly or implicitly. For example, our dating system is arguably implicitly indexical: it explicitly depends upon the identification of the birth of Christ, and our identification of that event perhaps depends in turn on features of our own position in time which, arguably, we can express only indexically (see Strawson 1959: Ch. 1, §2 and esp. p. 30).

I have no systematic argument for this claim of reducibility. It's just that, considering indexicals case by case ("today", "you", "that" are among those mentioned) one can see in each how to transform indexicality into anaphora in reporting speech. The thesis acquires a partial defence by showing that apparently unwanted consequences are either merely apparent or else not unwanted.

### **Action**

The canvassed reducibility is that anything that can be expressed by use of an indexical can also be reported, and so expressed, without using an indexical in the content-ascribing part of the report. This means that no thought essentially requires expression by means of an indexical; which in turn, by some standards, means that there are no indexical thoughts. Is this consistent with the special role of "indexical thought" in action?

It is unclear that the phenomena which are supposed to motivate the essential involvement of indexicality in action, or explanations thereof, really require that there are indexical thoughts in the envisaged sense. If I know that MS must make a call I may remain inactive if I don't realize that I am MS; whereas one can explain my making the call by attributing to me the thought that I must make a call. One might be tempted to infer that the thought that MS must make a call differs from the thought that I must make a call. Even if this inference is sound (which, in fact, I doubt), it does not connect with the reducibility thesis. That thesis requires, not that my knowledge be detachably expressible, but only that it be somehow non-indexically expressible. It is: MS's action is well explained by the hypothesis that he knows that he must make a call. A detailed example will bring out the point.

... suppose the commander says, "A hand grenade is thrown (tenseless) into this room on 1 December 1978." The soldiers will need to be able to judge whether 1 December 1978 is *today*, or years into the *past* or *future*. For without this information they will not know whether to take any action or to feel any urgency. (Sorabji 1983: 134)

The point is relevant to the claim that an indexical utterance has a self-standing non-indexical equivalent. But the present reducibility claim is weaker, and is not touched by Sorabji's point. If the soldiers rush out, we can explain their action by their realization that a grenade was to be thrown into the room *then*, where the "then" is not indexical, but anaphoric, dependent upon a

specification (implicit in the actual sentence I am using here) of the time of the reported realization. A test: were this token of “then” functioning as a genuine indexical (as other tokens of “then” may do), it could function without the time-fixing linguistic context; but evidently it could not. The implicit or explicit specification of the time of the soldiers’ realization is the token’s anaphoric antecedent.

If the kinds of thoughts which can properly be cited in explaining actions are non-indexically expressible, then indexicality cannot be essential to these explanations. The consequent appears correct: MS made the call because he knew he had to. Any plausible thesis of essential indexicality must be consistent with this fact. A weaker, and thus more plausible, thesis is this: a thought fit to explain an action is one which, were it to be expressed by the agent, would be expressed by means of at least one indexical. This is consistent with the thesis of reducibility. Indexicality is seen as a relation between a thinker and a thought, rather than a feature of the thought itself.

I read Perry as affirming only this weaker thesis (Perry 1979: 49). However, at earlier stages in his discussion one can discern some overstatements, e.g:

Imagine two lost campers who trust the same guidebook but disagree about where they are. If we were to try to characterize the beliefs of these campers without the use of indexicals, it would seem impossible to bring out the disagreement. (Perry, 1979: 35)

Perhaps each camper needs an indexical, or at least would most appropriately use one, to express his belief in a way which makes the dispute clear to his fellow. But *we* do not need to use one. We can say: They stood beside Gilmore Lake. John believed it was Eagle Lake but Bill believed it was Clyde Lake.

For Perry’s more finished formulation of the thesis, a distinction is required between a belief state and a belief (or belief content). People willing to sincerely assert the same sentence, e.g. “I am making a mess”, are thereby in the same belief state, but they may not share a belief. Sam believes that he is making a mess and Sally has the quite different belief that she is making a mess. But only Sam can believe what he believes by being in the state he shares with Sally, and it is this relation between belief state and belief that is essential to action (Perry 1979: 48–9). Perry thus concurs with the conclusion which seems forced upon us by reducibility: what matters is a relation between a thinker and a thought, rather than an intrinsic feature of the thought itself. On this approach, a proper recognition of the need for indexicality should have no tendency to promote any doctrines of limited or partial accessibility to thoughts.

When Perry comes to put the matter in a more theoretical perspective, he says that the phenomena he discusses make trouble for the “doctrine of propositions”. This doctrine sees propositions as objects of belief with fine-

grained content and permanent truth value. They cannot be identified through sentence-types containing indexicals, since these have different truth conditions on different occasions of utterance. So “there is a *missing conceptual ingredient*: a sense for which I am the reference” (Perry 1979: 37). Perry supposes that there should be a notion (“conceptual ingredient”) capable of explaining how content is related to utterances. For this purpose, a conceptual ingredient would have to be independently identifiable: it could not be merely whatever registers the relation between content and utterance. An alternative approach is to abandon explanatory pretensions at this point, and use a notion, say “concept” or “sense”, which serves merely to register the phenomena. In these terms, one would conclude from Perry’s discussion that an indexical token can express the same concept or sense as an anaphoric one but not as a proper name or a definite description. It is not that sense or concept would explain the phenomena; these notions would get their use from reporting it. Such an approach is discussed in more detail in §5 below.

### **Science**

Why has it seemed to some that indexicals should be banned from science? Supposedly, the answer is that indexicals introduce a perspective, whereas science is meant to be perspective-free. But if indexicals are reducible, i.e. can properly be reported without indexicality, we have to see the indexicality not as part of the content but as belonging only to how the content is presented. The claim that science must avoid indexicality would be an essentially stylistic recommendation, not touching the content of scientific theories. If a scientist were to present a theory using an indexical, we could report what he said, and thus the theory, without making use of an indexical. Whether or not to use indexicals would be a trivial dispute, of no philosophical interest.

An utterance of a scientific theory should be perspective-independent, in that the same theory should be available from arbitrary perspectives; but this does not entail that the utterance be perspective-free, made in a way that does not exploit any perspectival features.<sup>2</sup> Freedom from perspective is unmotivated, and probably impossible.

### **Time**

Without pretending to give anything like a serious exegesis of the complex dialectic involved in McTaggart’s argument for the unreality of time (1927),

2 Martin Davies suggested the terminological contrast “perspective-independent”, “perspective-free”, though I cannot vouch that he would accept the point I am here making by means of it. Further development of this line of thought might bear on various other matters, for example Jackson’s claim about what Mary knew.

I would like to mention one strand which relates closely to the present discussion.

Let us say that an A-series is a temporal series of events whose temporal features can be fully described in terms of the primitive vocabulary “past”, “present” and “future”, or expressions explicable in terms of these. We are to think of events as unextended in time. We have to do justice to two features, the unique location of every event and the passage of time. Allowing ourselves provisionally a mixture of the primitive and non-primitive vocabulary (i.e. quantification over moments) we might express these as follows:

- (1.0) *Unique location*: at any one moment, every event is just one of future, present and past.  
 (2.0) *Passage*: every event is both at some moment future, at some moment present, and at some moment past.

Striking out the non-primitive vocabulary yields a contradiction:

- (1.1) Every event is just one of future, present and past.  
 (2.1) Every event is both future and present and past.

We may try to remove the contradiction by replacing the quantification over moments in (1.0) and (2.0) in terms of the primitive vocabulary. Unique location is not too hard to express (at least in part) as follows: In the present, every event is just one of future, present and past. But with an eye to deriving a contradiction, we could regard the following as an equally good way of expressing uniqueness (perhaps entailed by the first way):

- (1.2) Every event is just one of: in the present, future, or in the present, present, or in the present, past.

Passage is more tricky. As an intermediate stage, still with some non-primitive vocabulary needing elimination, we might propose

For every event there is some moment, taken as present, for which the event is future, and some moment, taken as present, for which the event is present, and some moment, taken as present, for which the event is past.

A further round of replacing quantification over moments by the primitive vocabulary might produce

- (2.2) Every event is, in the present, future, and is, in the present, present, and is, in the present, past.

(1.2) and (2.2) contradict. Within any A-series, the demand of unique location contradicts that of passage. Hence there is no A-series.

I think the proposals under consideration here can make a small contribution to understanding. “Present” and the rest have, like the personal pronouns, both an indexical and an anaphoric role and the indexical role is reducible. Your utterance of “Our finest hour lies in the future” exploits what can properly be regarded as indexicality: a non-linguistic feature, the time of your utterance, is required as an index to determine the contribution which “in the future” makes to its truth conditions. However, if I report this as your having said that our finest hour lies in the future, the same phrase is anaphoric, not indexical: its contribution to truth conditions is not a function of the time of my report, which might occur later than our finest hour, but is determined by the implicit specification of the time of your utterance, effected by my use of the past tense of “said”. (I gloss over the complications involved in “sequence of tenses”. For example, if my report occurs after the time of the event you predict, I should use the past tense in the content-specifying part of my report: You said that our finest hour *lay* in the future.) These uses are closely related and complementary, and, I propose, one could not coherently require that these expressions be taken as primitive vocabulary without allowing both their uses. (An argument for the general claim, not applied explicitly to “future” etc., that demonstrative uses presuppose anaphoric ones is given by Brandom 1994: see e.g. 464–5.) But once the inseparability of the uses is acknowledged, one has to acknowledge that some expressions fit to be antecedents to the anaphoric uses must be included within the primitive vocabulary. If one takes tenses, as the most common, then we can state the thesis of passage in such a way that it does not conflict with the thesis of unique location:

(2.3) Every event is, was or will be future, and is, was or will be present, and is, was or will be past.

The thesis of unique location ensures that each event which satisfies a conjunct will do so in virtue of satisfying just one of the disjunction of tenses. The theses are complementary rather than conflicting.

### **God**

If there are no indexical thoughts (as discussed under the headings “Reducibility” and “Action” above), there is no indexical knowledge: no knowledge expressible only by means of an indexical. This gives a quick answer to an ancient argument, which could be phrased as follows: The use of a temporal indexical requires that one be in time. Hence a timeless God cannot have temporally indexical knowledge. Hence there is something a timeless God could not know. If the reducibility thesis is accepted, however,

“indexical knowledge” can be non-indexically expressed, so the argument gives no good reason for supposing it to be unavailable to a timeless God. Following our guide, we will individuate things known by the test of speech reports: if I report you as having said something, and I know the something, then I know what you know (“the same *thing* as you”). If Paul utters “Now I see the light”, a timeless God can report him as having seen the light then, and so can know that he saw the light then, and so, without using an indexical, can know the thing which Paul knew.

This is not the end of the story, for a related argument remains to be addressed. Can an eternal God refer to moments of time, or specific events in time? If he cannot, then he cannot know what his creatures know, for he cannot so much as report what they say when they express their knowledge, since he cannot identify the events of utterance and so cannot produce the scene-setting part of the report. This argument, whatever its merit, goes well beyond anything specially related to indexicality, for it casts doubt quite generally on an eternal being’s capacity to identify things in time; so it lies beyond my present purview.

## 5 The Fregean connection

Although Frege has so far hardly been mentioned, my approach (and examples) are based on some aspects of his approach. But instead of trying to make anything of “modes of presentation” as a basis for sense, I have tried to make something of the accuracy of reported speech. The justification, in Fregean terms, is that the content-ascribing words in a report of speech ought to match the originals in sense. (Strictly, we should be able to infer only that the customary sense of the content-ascribing words matches the customary sense of the originals, and we cannot go on to infer that the sense they actually have in their indirect context matches that of the originals. But I take for granted Dummett’s modification, according to which the sense/reference distinction evaporates in indirect contexts: indirect reference = indirect sense = customary sense.) Let us use “sense” for Fregean senses as individuated by modes of presentation, in turn regarded as capable of being individuated independently of the needs of semantics; and “meaning” for something similar to Fregean senses but individuated by a combination of the demands of reported speech and a Fregean test in terms of rational cotenability. The question for this section is how senses and meanings compare for grain; the conclusion is that senses are in some respects finer-grained and in some respects coarser-grained than meanings. In each case, I think meanings have the grain more appropriate to semantic taxonomy, though I do not argue for this.

The project as stated cannot be undertaken with full seriousness, for Frege never tells us how to individuate modes of presentation. I will assume in this part of the discussion that modes of presentation are, at least paradigmatically,

perceptual. This assumption is not really justified by Frege's text, though many readers of Frege appear to take the text this way. My discussion will suggest that there is no independent account of modes which provides the right taxonomy for semantics. Fregeans do not have to see this as essentially anti-Fregean (indeed, I do not intend it that way); rather, they might see the claim as a point in favour of allowing modes of presentation either a merely heuristic role, or a dependent one: either functioning as a striking example (in the triangle case, for instance), not to be generalized, or as having a nature which is to be fixed by the demands of semantic theory, rather than being an independent input to such theory.

Just as any two things are similar in some respects and dissimilar in others, just about any two perceptions of an object can be counted as cases in which the object is presented under a common mode (e.g. the perceptual), and as cases in which it is presented under distinct modes. This means that there would be a glaring gap in Frege's account of sense, if mode of presentation is supposed to play an independent role. The gap would also make it difficult to undertake the task of this section. I circumvent the difficulty as follows: in the first kind of examples to be discussed, which relate to the first person, we have a specific Fregean pronouncement on the individuation of modes of presentation and hence, on the present assumption, of senses. In the second case, in which our considerations require acknowledgement of different meanings, it would be very hard to discern any basis for distinctness of mode of presentation, and hence of sense.

Frege claimed that "Everyone is presented to himself in a special and primitive way, in which he is presented to no-one else" (1918: 359). If an expression's sense is fixed by such a mode of presentation of a person, then it can be grasped only by that person. Readers of Frege often take it that he is claiming that the first person pronoun is such an expression. Although I do not share this reading,<sup>3</sup> it would have the consequence that we have difference of sense yet sameness of meaning. Since, as the example at the start of the essay suggested, I can say by uttering "I am a fool" what you say when you utter the words "You are a fool", the present methodology dictates that a token of "I" can have the same meaning as a token of "you". So, granted all

3 It is not clear that Frege himself thought that the sense of any expression is so fixed. Frege's commitment to communicable "I" thoughts appears not only in explicit form, when he offers the rather hasty suggestion about the sense which can be grasped by others ("he who is speaking to you at this moment", p. 360), but also, more interestingly, in an implicit way, when, having earlier reported Dr Lauben as having uttered the words "I was wounded", he slips with silken ease into reporting Dr Lauben as having the thought that *he* was wounded (p. 359). If the report is accurate, as indeed it seems, then Frege is committed to there also being a mode of presentation of a subject which others can have, where this mode constitutes the sense of that subject's tokens of the first person pronoun, and is shared by appropriate tokens of the second and third person pronouns in the mouth of others.

the assumptions, senses are finer-grained than meanings, though in a way that is far from counting in their favour.

I stipulate that the meaning of a complete sentence is the thought it expresses, and that a sufficient condition for distinctness of thoughts can be expressed in the Fregean terms of rational cotenability. One version of the criterion is this: thoughts differ if either is rationally cotenable with the negation of the other. The criterion delivers, I believe, that there are cases of indexical tokens *a* and *b*, of the same type and uttered in what by many reasonable standards counts as the same context, for which the thoughts expressed by *Fa* and *Fb* differ, so that the tokens themselves differ in meaning; by our guiding hypothesis, the same will go for their corresponding anaphors in reports. In Perry's famous example (1977), slightly modified, a person seeing a ship out of a window utters the words "That was built in Japan" and a moment later utters the words "That was not built in Japan". As interpreters, the example continues, we must treat the reference of "that" as the same on each occasion. Yet various things might conspire to make the speaker believe without irrationality that he had referred to different ships, and had expressed two truths. Applying Frege's criterion of difference, we conclude that the thought expressed by the second utterance is not the negation of that expressed by the first, and, given the compositionality of meaning, the only available explanation would appear to lie in different meanings of "that". Yet there seems no prospect of identifying a difference of mode of presentation on any independent basis. It is not plausible to say in general that the passage of time, or an influx of information concerning a presented scene, changes its mode of presentation: that would make it impossible to sustain a thought over time and over informational enrichment. One might make a special case for the susceptibility of "that" to shifting modes of presentation. If so, my general point is best made by considering how, according to the methodology of the paper, a report of the envisaged speaker should be understood.

The report could go: Seeing a ship through a window, he said that it was built in Japan, and seeing the same ship through the same window a moment later, he said that it was not built in Japan. This attribution does not imply irrationality in the sincere speaker whose words are thus reported. (This is consistent with it being more often than not the case that such a speaker is, in fact, irrational.) This means that the occurrences of the anaphoric "it" must differ in meaning, even if they anaphorically depend upon the same words and the same context. It would be hopeless to try to associate, on independent grounds, distinct modes of presentation with the distinct occurrences of "it": the reporter may not have perceived the ship, and need not know how it looked to the person whose speech he correctly reports. (This kind of example can be developed as an objection to thinking of demonstrative pronouns as free variables.) In these examples, meanings are finer-grained than senses.

My position is supposed to be Fregean, except for three points: I put tokens



at the centre of the subject matter; I remove mode of presentation from a central role in the explication of sense, allowing some of the work it was fashioned for to be performed by constraints on reporting speech; and although these constraints help found an equivalence relation, I find no need to think of senses as entities.<sup>4</sup>

4 I am very grateful for the comments I received as a result of the Symposium, notably Jimmy Altham's illuminating response and written comments from Adam Morton. Altham's comments show that the guiding hypothesis requires considerably more amplification than I have been able to give it here. Among the contributions from the audience at the Symposium, I particularly remember those by Bob Hale, James Higginbotham, O. A. Ladimeji, Philip Percival and Timothy Williamson, all of which led to changes in the text. Thanks also to Maite Ezcurdia, Christopher Hughes, Keith Hossack, Michael Martin, Gabriel Segal, Richard Sorabji, Mark Textor and David Wiggins for many valuable discussions of these topics. One idea in this paper overlaps with one in Ezcurdia (1996); we are not clear who, if either, thought of it first.

# Names, fictional names, and “really”

---

## I Introduction

Fictional names pose at least two kinds of problem: how, if at all, do they mean anything? And how, if at all, do they affect the truth or other semantic value of sentences in which they occur? Answering the first question is especially difficult if one holds both of the two opinions (1) that all genuine names have bearers and (2) that fictional names do not. One must then conclude that fictional names are a kind of non-name, and one will have to try to explain away their apparent intelligibility. In this paper I shall take for granted one of the views that lead to this difficulty, namely that many fictional names lack bearers. However, I shall take issue with the other component, that a name must have a bearer. A full treatment requires two parts. First, one should undermine the arguments designed to show that all names (or all names of a certain kind) must have bearers. Then one should provide an alternative account. The first of these tasks is undertaken in §2. The second task is a large one and only a small part of it will be attempted here, starting in §3.<sup>1</sup> There I consider whether we can make use of the notion of make-believe to give a positive account of what is involved in understanding a fictional name. In §4 I discuss what difference, if any, “really” makes to the truth conditions of sentences, especially those which involve fiction.

## 2 Evans’s arguments for the Russellian view of names

The idea that names must have bearers has a long history, going back at least to the middle ages, for example, to Anselm:

1 The task is taken further in essay XII below.

Si vero non significat aliquid, non est nomen. (quoted by Henry 1984: 3)

The view is helped on by the notion of “signification”: an expression without signification, as an empty name might be called, would be one without significance, that is, without meaning. Nearer our own times, Russell is a famous proponent:

What does not name anything is not a name and therefore if intended to be a name is a symbol devoid of meaning. (1919: 179)<sup>2</sup>

Gareth Evans has adopted essentially this view, claiming that at least a large class of ordinary proper names are “Russellian” in that to suppose they have no bearer is to suppose that there are no truth conditions for sentences in which they are used. We can call contrasting views “Fregean”: these allow that there can be genuine names, capable of being understood, which do not have a bearer. (The label “Fregean” does not imply anything more than this; plenty of “Fregean” views would not have been held by Frege.)

A surprising number of the features Evans associates with Russellian singular terms can be accepted by a Fregean. It can be agreed that the use of a name requires discriminating knowledge of its bearer, if it has one; that names are normally used information-invokingly; and that one can properly attribute a meaning to a putative name only if there is, for each sentence in which it occurs, a truth knowledge of which would suffice for understanding. I will first show how these points can be met; and then move to Evans’s two official arguments for the claim that singular terms which are standardly used information-invokingly are Russellian, and so require bearers.

## 2.1 *Discriminating knowledge*

Evans says that Russellian singular terms are governed by Russell’s principle, which at one point he expresses as follows: “the subject must have a capacity to distinguish the object of his judgement from all other things” (1982: 89). Perhaps no version of this principle should be accepted. However, this issue is independent of the debate between Russellian and Fregean accounts of names, since theorists of both kinds can accept the principle. The sentence from Evans just quoted must really identify a capacity that comes into play only when a subject’s judgement has an object, which not all judgements do: there are purely general judgements like “there exists at least one dog”. Once the sentence has been appropriately generalized, it is acceptable to a Fregean;

2 Cf. “the linguistic function of a proper name is completely exhausted by the fact that it names its bearer” (Kripke 1979: 240).

for a Fregean can accept that a subject must have a capacity to distinguish any object of his judgement from all other things.

A Fregean could even accept what Evans says is an *excessively demanding* version of the principle, as applied by Dummett, on Frege’s behalf:

The general idea was that a grasp of the sense of the name consisted in a capacity to say, of any given object, whether it was the referent or bearer of the name. (Dummett 1973: 488)

The natural reading of this principle is that persons who understand “*a*” should be able to answer arbitrary questions of the form “Is this *a*?”. Even for empty names, these questions are perfectly intelligible, by Fregean standards (though the answer for such names ought always to be “No”). Such a theorist can impose, or refrain from imposing, a condition of this kind, as he sees fit in the light of independent considerations.

## 2.2 Invoking information

Understanding an expression consists in immersion in the practice of its use. Evans’s notion of the information-invoking use of a name can be seen as an attempt to characterize this immersion for this class of expressions. According to Evans, a name (more generally, a singular term) is used in an information-invoking way if the hearer of an utterance containing it must, in order to understand the utterance, link it with some information in his possession (cf. Evans 1982: 305). (This must not slide into the stronger condition that some information must be linked with it.) Evans allows, what certainly should be allowed, that an informational state apparently about an object may not be about any object at all: “An informational state may be of *nothing*: this will be the case if no object served as input to the informational system when the information was produced” (p. 128).<sup>3</sup> Moreover, he allows that identity of an informational state can be defined in a way that does not require a common object as its source: “When two states embody *the same information*, they are necessarily such that if the one is of an object *x*, then so is the other” (p. 129). The requirement is merely that the information be from the same object if from any. So there are no general reasons why a Fregean should not accept that “the use of an ordinary proper name is always information-invoking” (p. 310).

3 For this to be true, “input” must be regarded as a technical term, which awaits clarification.

### 2.3 Understanding and knowledge

For any utterance that can be understood, Evans claims, “there is some true proposition such that knowledge of its truth constitutes understanding the utterance” (p. 330). Although it is not clear that Frege himself would accept this condition, one who is Fregean as I use the term here can allow for such truths, ones derivable from claims like: “Vulcan” refers to something iff that thing is Vulcan.<sup>4</sup> Evans thus overshoots when he claims: “But when the singular term is information-invoking, it seems to me that if there is nothing to which it refers, then we must deny that there is any such true proposition” (p. 330).

Evans makes specific claims about what understanding an information-invoking expression requires, for example: “if the conditions of understanding require the audience to be thinking of an object which is also the object the speaker refers to, then, if there is no object the speaker refers to, those conditions cannot be satisfied, and his remark cannot be understood” (p. 332). If we were to replace “an object which is also the object” by “whatever object”, this conditional would have an antecedent acceptable to the Fregean, but the conditional as a whole would be false. In subsequent subsections I muster what arguments I can find in Evans that would support his version of the antecedent rather than an existentially neutral one.

### 2.4 The first argument: truth is seamless

This first argument starts with the lemma that in order to understand singular terms which are intended to invoke identifying information “one must oneself believe there is something to which the term refers” (p. 326). Since this belief is in part constitutive of understanding, and since understanding is a form of knowledge, the belief must also be knowledge, and so true. “Truth is seamless: there can be no truth which it requires the acceptance of a falsehood to appreciate” (p. 331).

One could present the argument as follows:

4 Here I assume the kind of reference-condition semantics described by Burge (1974), in which names are treated by axioms like

$$\forall x (\text{“Hesperus” refers to } x \text{ iff } x = \text{Hesperus}).$$

The setting is negative free logic: atoms with empty names are false, and universal and existential quantifier rules are modified. An axiom of the above form for an empty name like “Vulcan” is true: the right hand side is false for each value of  $x$ , leading to the appropriate verdict that there is nothing to which “Vulcan” refers. One interesting feature of the theory is that the semantics associates names neither with an object nor with a description (in the usual qualitative sense of “description”).

- (1) Understanding an utterance containing an information-invoking use of a singular term requires believing that there is something the term refers to.
- (2) Understanding is knowing what is said.
- (3) “Truth is seamless”: knowing what is said cannot require having a false belief.
- (4) Hence the belief specified in (1) must be true: there is something the term refers to.

The two most dubious premises, which are therefore natural targets for the Fregean, are (1) and (3). The latter is a hasty, and unsatisfactory, attempt to formulate something like a “no false lemmas” account of knowledge. It is unsatisfactory, because I can know that I have at least one false belief (who doesn’t?), but this is a piece of knowledge I could not have unless I believe a falsehood.<sup>5</sup> One option, therefore, would be to develop an improved version of this constraint on knowledge, and see whether using it instead of (3) would affect the soundness of Evans’s argument. Whatever the upshot of this approach, the Fregean would still wish to challenge (1); and I believe that Evans has, at this stage of his account, provided no reason at all to believe it.

There is little doubt that in many cases of communication a hearer expects the names used by a speaker to refer, and takes it that he needs to know to which objects they refer. The question is the extent to which this expectation is a necessary product of understanding. I will follow Evans in setting aside those cases, like fiction and negative existentials, in which it is clear that understanding takes place when this expectation is absent – conniving uses, as Evans calls them.

The Fregean can grant that something very close to this expectation belongs to understanding. Evans allows that “the notion of the intended referent is rather like the notion of a *target*” (p. 317), something that might be aimed at yet not hit, and something that might figure in different ways at different levels of action-plans (aiming at that man and aiming at the man who insulted you may not amount to the same thing, even if that man is the one who insulted you). The Fregean can accept the following:

one who understands a speaker using a name must be able to answer the question: what did the speaker aim to refer to?

Answering the question correctly does not commit the understander to belief in the existence of such an object. If you say that Vulcan is heavier than Neptune, I know that you aimed to refer to Vulcan. There is no inference to: there is something to which you aimed to refer. On the contrary, if I know that Vulcan does not exist, I will know that you failed in your aim.

5 Thanks to Peter Klein for pointing this out.

This comports with Evans’s claim that “for a hearer to understand a speaker . . . , he must know which object the speaker intends to refer to” (p. 318). The ellipsis contains the words: “making a reference”; the Fregean could replace them by “using a referring expression”. What Evans needs, but does not in this phase of the argument provide, is an argument for an understanding-condition which requires the hearer not merely to have a belief about what the speaker aimed to refer to, but to believe that the speaker succeeded in his referential aim.

The upshot is that even if hearers typically expect names to have referents, we have not yet been given a reason to suppose that this is part of what is involved in understanding them. Understanding requires appreciating what the speaker is aiming to refer to, an appreciation which can amount to knowledge even when the aim radically misfires through lack of a referent. Fregeans can properly reject Evans’s lemma (1). In any case, the lemma would create a fracture in our name-using practices between what Evans calls the conniving uses (involved in fiction and sincere negative existentials) and normal uses, for the former uses, even in assertion, let alone in other speech acts, clearly do not require, on the part of either speaker or hearer, belief that the singular terms refer. The Fregean, by contrast, sees unity: reference is needed for atomic truth, and so must be believed in by a sincere speaker and an accepting hearer in such cases; but it is not needed for meaning or understanding, and what is needed can be the same whether or not the truth of an atom is what is in question.

A lively name-using practice may make room for an existential question. Before doubts occur, we use the name believing the associated information. As doubts creep in, our confidence in the associated information diminishes. It seems clear that this diminishing confidence, the terminus of which is doubting all of it, is to be sharply distinguished from gradual loss of understanding. Indeed, if the doubts culminate in the belief that the supposed object does not exist, so that uses of the name come to be classified as conniving, it seems we cannot allow that loss of understanding occurs, else the belief in non-existence would not be the negation of the earlier belief in existence.

## **2.5 The second argument: the demand of unity**

. . . the challenge, to those who wish to argue that information-invoking referential communication can take place in the absence of an object, is to state a communication-inducing relation between the origin of the speaker’s information and the origin of the hearer’s information which does not presuppose that the information originates in episodes involving the same object, but which, when the information *is* from an object, holds in just those cases when it is from the same object. I do not say that it cannot be done, but I myself do not see how to do it. (p. 337)

Here is a proposal to meet the challenge:

the information invoked by the speaker must have the same source or origin as that invoked by the hearer.

Information originating (in the appropriate way) in the same object has the same source; so does information originating (in the appropriate way) in the same witness. Evans explicitly mentions what he calls the “journalistic sense” of the word “source” (p. 337), and he uses it in such a way as to exclude what is needed for the displayed principle: in the journalistic sense, Deep Throat, rather than Nixon, is the source of Woodward and Bernstein’s information. However, “source” need not be so restricted, or, if it need, let us choose another word, say “origin”. The origin of a malicious rumour may be an evil tongue; the origin of information about the centre of the earth might be a complex theory, or alternatively might be signals emitted by an instrument located at that spot, in which case the information originates at the earth’s centre. There is a reasonably clear and non-disjunctive notion of the origin of information, and it is surprising that it seems not have struck Evans as a candidate solution to his problem of unity.<sup>6</sup> I will test this proposal against his detailed arguments.

An early example is a demonstrative case where communication takes place “between a speaker and a hearer who are sufficiently far away from one another to observe different parts of the same rather large object” (p. 333). Evans envisages that the only alternative candidate, other than the object itself, for bringing out what speaker and hearer share is some spatial descriptions; but these would be different as between speaker and hearer, so not shared after all. “The only thing that can bring unity out of this diversity is the fact that there exists an object of which both spatial descriptions are true” (p. 334). However, there is an alternative: the bodies of information each uses to guide their speech or understanding have the same origin. In this case, that is because the information comes from the same object; but the condition is general and will work when this is not so.

Turning explicitly to the case in which there is supposed to be understanding in the absence of an object, Evans writes:

The only candidate communication-allowing relation between the thoughts of speaker and hearer, which is discoverable in the absence of an object – that they both involve exactly the same way of identifying, or

6 Admittedly, the notion is a hard one to explicate. If a fit of anger makes a person concoct a falsehood involving a fictitious event, neither the anger nor the person is the reference of a term for the event, and the person not the anger counts as the “origin” of the information. Some other complexities are mentioned later in the present essay and are discussed in essay XII below.



purporting to identify, an object – is far too strong a requirement to impose upon referential communication in general. (p. 336)

One can agree with Evans when he says that identity of modes of identification is too strong a condition to work in general, yet disagree that this condition is the only one available for the case in which there is no object. On the contrary, the present proposal meets the case with a weaker condition, for distinct bodies of information may have a common origin.

The notion of origin needs attention. An over-imaginative, or self-deceiving, or evil tongue,  $T_0$ , may start a rumour which is embellished by others. The rumour is that there’s a dragon, Fiamma, who lives in the mountain just south of the village and whose preferred diet is human babies. You hear the rumour from  $T_1$  and I hear it from  $T_2$ , each of whom heard it, on separate occasions, from  $T_0$ . You say that Fiamma is green, trusting to  $T_1$ ’s embellishment, and I say she is red, trusting to  $T_2$ ’s. By some standard, our Fiamma-related information has different origins; but there is a standard which rules that these different bodies of information have the same origin, in  $T_0$ . This is the standard we need: it correctly represents us as disagreeing about Fiamma’s colour. It rules as it does because although the information that Fiamma is green is new to  $T_1$ , and the information that Fiamma is red is new to  $T_2$ , both these pieces of information were intended by their producers to link to Fiamma, so the “ultimate” origin lies further back, with  $T_0$ . This is the origin that is invoked by my proposal.<sup>7</sup>

There can be equivalent bodies of information with different origins, in which case we have to see them as involving distinct name-using practices.  $T_0$  lives on the north side of the mountain. North siders have no contacts with south siders. Quite coincidentally, a south sider spreads a rumour about a dragon, Fiamma, who lives in the mountain just north of the village and whose preferred diet is human babies. Were a south sider to meet a north sider and begin a “Fiamma”-involving conversation, there would be an illusion of understanding; but no more than an illusion since their information has no common origin. To put it in a tendentious way, though one whose naturalness is of interest, they would be talking about different dragons, whether or not they could ever know this. This kind of case mirrors for empty names things that Evans has taught us about non-empty names.

Evans was well aware of the point of view I have been expressing:

<sup>7</sup> Evans makes a similar point in connection with non-empty names: a hearer’s information may originate in “quite different episodes” (p. 337) from those in which the speaker’s information originated. The problem of giving a proper account of origin is barely scratched by the proposal in the text, though a developed Fregeanism essentially requires such an account.

When there is this common information, it seems very tempting to suppose that we can draw just the same distinction between understanding and misunderstanding as we do in the case where there is a referent. We cannot say that the hearer who understands has got hold of the right object; but he can, so to speak, attach the remark to the right information (p. 339).

He says that he has tried to weaken this temptation, by arguing that the notion of the “right” information cannot be made to work properly (this alludes to the unity argument) and that understanding a remark cannot involve a false belief (this alludes to the “truth is seamless” argument). I have tried to show that both these arguments fail. I now turn to what is involved in understanding empty names, or rather, to understanding a subgroup of these: fictional names.

### 3 Understanding fiction

Let us use “quasi-understanding” to denote the state people are in with respect to fictional names, and sentences containing them, which, as we would ordinarily say, they understand. Given Evans’s Russellian view of names, together with his disinclination to believe that fictional characters really exist, he is forced to conclude that quasi-understanding is not a species of understanding, for fictional names are not a species of name. Although this is not to be accepted, we can none the less explore whether a Fregean can use Evans’s account of quasi-understanding to explain what it is to engage in a practice of using a fictional name. The conclusion of this section is that although it is of interest that some name-using practices exist only in virtue of the imaginative production of, and engagement with, fiction, it is not the case that every participant in such a practice needs to appreciate the role that fiction plays in it.<sup>8</sup>

Before amplifying the remark that Evans’s view is not to be accepted, I briefly describe Walton’s notion of quasi-fear, upon which Evans explicitly models his account of quasi-understanding.

#### 3.1 *Quasi-fear*

Walton asks us to consider Charles, who has the bodily signs of fear while watching a horror movie. Let us use “quasi-fear” with initial neutrality, for the state, whatever it is, we are inclined to call fear in Charles, made manifest

8 This conclusion is argued for below (§3.2); it will already have been accepted by those convinced by the “Shylock” example in §3.3 in essay XII below.

by the fear-like bodily symptoms.<sup>9</sup> Walton argues that Charles’s quasi-fear is not fear. The argument involves a positive and then a negative phase. The positive phase is designed to establish that it is make-believedly the case that Charles is afraid. In watching the movie Charles is prompted to imagine a situation – that the deadly green slime is closing in on him – in which he is in danger. This extends the make-believe of the movie, so that it is make-believedly the case that Charles is aware that he is in danger. A further extension makes it make-believedly the case that Charles is afraid of the slime, an extension grounded in the thought that if things were as Charles imagines with the slime, that is, were the slime evidently posing a danger to him, he would be afraid of it. Since it might be both make-believedly and really true that Charles is afraid, the argument requires a negative phase. This consists in arguing that Charles cannot really be afraid since he does not really take himself to be in danger.<sup>10</sup> Putting the two parts together, we have an account of Charles’s quasi-fear as involving it being make-believedly the case that he is afraid though he is not really afraid.

In the case of quasi-understanding, the positive phase is supposedly achieved by reflecting on how understanding needs imaginative engagement with the fiction; the negative phase is supplied by the doctrine that names are Russellian, and so cannot be genuinely understood.

In our present dialectical position, we have no reason to accept the negative phase of the argument about quasi-understanding. Quasi-understanding is very different from quasi-fear. There is something initially puzzling about saying that Charles is afraid. We are inclined to say: he’s not *really* afraid. Questions such as: “how can he be afraid given that he knows the slime isn’t real, so that he’s not in danger?” and “why doesn’t he just turn the TV off?” and “why does he watch if it affects him like that?” are genuine questions, not mere philosophers’ questions. By contrast, there is nothing initially puzzling about a person’s understanding of a fictional sentence (a high proportion of sentences to which infants are exposed fall into this category). We have no inclination to ask questions like “how can he understand, given that he knows it’s just a story?” or “given that he knows there’s no such person as Hamlet?”. On the face of it, understanding fictional sentences can serve as a paradigm of genuine linguistic understanding, whereas fiction-induced (quasi)-fear cannot serve as a paradigm of genuine fear. The upshot is that when we discover an inconsistency between the natural claim that quasi-understanding is

9 “His muscles are tensed, he clutches his chair, his pulse quickens, his adrenaline flows. Let us call this physiological/psychological state ‘quasi-fear’. Whether it is actual fear (or a component of actual fear) is the question at issue.” (Walton 1978: 6)

10 The argument is not as simple as this, for Walton takes account of those who (irrationally) seem to fear something they say they recognize to be harmless. Such people are motivated by fear, whereas Charles is not (cf. Walton 1990: 201–2).

a species of understanding and the claim that a name (of the relevant kind) cannot be understood unless it has a bearer, we will need some persuading that the resolution of the tension requires giving up the natural claim. The thesis that quasi-fear is not fear explains something puzzling; the thesis that quasi-understanding is not understanding creates a puzzle where before nothing was puzzling.

We can turn our backs, then, on the negative phase of the argument about quasi-understanding: quasi-understanding is a species of understanding. There remains the question of what species. It does indeed seem that (quasi)-understanding fictional names typically involves some imaginative engagement with the fiction, rather as becoming quasi-afraid does; so it may be that a full account of the practices of using fictional names will allude to this engagement. In this spirit, we can ask whether the positive phase of Evans’s argument about quasi-understanding can offer insights available to Fregean theorists of names. I include under this umbrella those who hold that an expression having sense but lacking reference can make a contribution to the truth-conditions of sentences in which it occurs, and also those who, unlike Frege, hold that such a sentence whose truth conditions do not obtain is false.

### **3.2 Make-believe and understanding**

We can agree with Evans that “Understanding these [fictional] uses of singular terms requires from the hearer something of *the same general kind* as is required to understand ordinary information-invoking uses of singular terms: the hearer must possess some information or misinformation, and somehow bring it to bear upon his interpretation of the remark” (Evans 1982: 344). There is room for dispute about the details.

The kind of fiction that is relevant to the discussion is the kind Evans calls “existentially creative”, involving the pretence that something exists which does not really exist (as opposed to fictions in which there is pretence, concerning things which exist, that they are other than they really are). Evans’s attractive idea is that existentially creative fiction makes room for fictional names: our capacity to use them systematically depends upon our engaging in an existentially creative fiction. Translating this into a condition on understanding would suggest something like: a name is capable of being understood if some fiction has it that it has a bearer. This is analogous to a sufficient condition for the possibility of understanding ordinary names: such a name is capable of being understood if it has a bearer. One might expect that structurally similar embellishments of each condition would yield something sufficient for actual understanding and not just the possibility thereof. These conditions would then feed in to the understanding of whole sentences in standard compositional ways.

Using “F( )” to mean “the fiction has it that ( )”, a preliminary interpretation

sees Evans as saying that existential creativity makes room for sentences  $s$  such that  $F(s \text{ has truth conditions})$ , the supposed corollary being that  $s$  is capable of being quasi-understood. We can express this as the hypothesis:

(5) If  $F(s \text{ has truth conditions})$  then  $s$  is capable of being quasi-understood.

Suppose the story teller says: “And then Alice came to the dragon’s cave, and said the magic formula the friendly genie had given her, and which no human ear could understand: ‘erty uiop asd’. The dragon heard and realized that he was not to attack Alice.” This ensures that  $F(\text{“erty uiop asd” has truth conditions})$  yet the quoted string is not capable of being quasi-understood by the story’s intended audience; so (5) must be rejected. The example is quite unlike the sentence “Hamlet was indecisive”, though it is part of the pretence that it is capable of being understood (at least by dragons). We therefore need to look for a more nuanced account of the way in which a fiction can yield a truth condition.

One emerges from Evans’s discussion of fiction with a vivid appreciation of the fact that there is an important structural similarity between thinking within and about fiction on the one hand, and thinking within and about reality on the other. Anything one can really do with language one can also make-believe to do. Suppose one thought that, for factual discourse, it was approximately correct to say (with appropriate qualifications) that understanding a sentence  $s$  (which in fact means that  $p$ ) is knowing that  $s$  is true iff  $p$ . A direct application of the point about structural similarity would suggest the following:

(6) One understands a sentence  $s$  which has a use only in fiction iff one knows that  $F(s \text{ is true iff } p)$ .

However, while knowing that  $F(s \text{ is true iff } p)$  may be a sufficient condition for understanding  $s$ , it seems to me not to be a necessary one. Suppose someone reads a novel as a historical narrative (some novels are written so as to encourage such a reading). I think he has none the less understood it, one sign being that he may need to exploit this understanding to appreciate what we, trying to disabuse him of his error, are saying. We insist: “But Anna Karenina isn’t a real person; she’s only a character in a novel”, and the too credulous reader needs to bring to bear an existing understanding of the name in order to appreciate what we have said.

A Fregean who accepts a Burge-style negative free logic can insist that T-sentences for fictional sentences are factually true. It is not just that, in the fiction, “Anna had grey eyes” is true iff Anna had grey eyes; this can be read as true absolutely. On this reading, the right hand side of the biconditional, containing an atom with an empty name, is false; hence so too is the left hand side. (We can also make room for other readings in which some or all

parts of the biconditional are implicitly prefixed with an “F” operator, or are assessed for truth-in-fiction as opposed to truth.) Called upon to explain how “Anna” has a meaning in the metalanguage, we can indeed allude to Tolstoy’s work, allowing that it would not have had this meaning had he never written. But we have no need to suppose that participants in the practice which exists only thanks to his work need appreciate this fact: there might be benightedly over-credulous ones. Such participants will bring information to bear in understanding the name (for example, that Anna was wooed by Vronsky); but they do not need to know that this is misinformation, nor that it is the special kind of misinformation which is created by novelists. So we can agree with Evans that the intelligibility of fictional names depends upon existentially creative make-believe, without advancing to the conclusion that the make-believe needs to be known to everyone who understands such a name.

It may be that active imaginative engagement of a kind the subject must realize is appropriate only to a fiction is required for some aspects of the appreciation of fiction and other works of art (as Walton (1990) makes highly plausible); but this is not to say that this sort of engagement is required for understanding fictional sentences.

The structural similarity between real and fictional linguistic activities runs deeper than was at first anticipated: participation in a fiction-based name-using practice may be indistinguishable to the participant from participation in a real-object-based practice. This upshot is only to be expected if, as seems to me correct, fiction and error are species of a single genus, for it is obvious that one can participate in a practice based upon error without appreciating this basis.

#### 4 “Really”

We move back and forth between fiction and reality in various ways. We exploit fiction while commenting on its merely fictional nature; we think about and have attitudes towards fictional characters and we compare them with real ones; encountering fiction produces real effects, apparently of fear or joy (or, at least, of raised pulse rate). Does the expression of such activities require (or at least make natural) a truth-condition affecting operator?

An interesting contemporary view, pioneered by Evans and developed by Wiggins, is that the answer is affirmative, and that “really” is precisely such an operator. Both philosophers use this notion to explain the possibility of true singular negative existentials. The idea is that “really” occurs in the logical form of, for example, “Hamlet does not exist”, and a proper appreciation of how this is so is essential to a correct account of how such sentences can be true. The aim of this section is to make an opposed suggestion: as used in ordinary English, “really” performs no such function; nor is there any need to introduce a technical notion which does.

Evans clearly recognized that someone who, like himself, adopts a generally Russellian approach to names owes an account of true negative existentials which have a name in the position of grammatical subject, for these otherwise stand as plain counterexamples to the Russellian view. There are two problems. First, on the Russellian view, such sentences would appear to be meaningless, for they appear to contain a name lacking a bearer. Second, once this difficulty has somehow been overcome, it remains to explain how such sentences can be true.

We have already seen how Evans addresses an aspect of the first issue, at least for sentences like “Hamlet does not exist”: he denies that this fictional name is really a name, or that it can be understood. But this makes the second problem even harder; for how can an expression which cannot be genuinely understood be used in a sentence which is genuinely true, and so can be genuinely understood? His account of “really” addresses this second problem.

The suggestion is that true negative existentials function rather like a “move within a pretence [which expresses] the fact that it is a pretence” (1982: 369). To use an image from Wiggins (1995), a sincere user of such a sentence is seen as standing with one foot in the fiction, where the singular term leads a simulacrum of a normal life, and one foot in the real world, from which perspective the term can be used to state the fact which makes its behaviour abnormal.

Evans implemented this idea by suggesting that we can think of a sentence like “Hamlet does not exist” as equivalent to “Hamlet does not really exist”, whose form is more fully described by:

Not (really (Hamlet exists)).

Here the contained “Hamlet exists” is a sentence which represents a move within a pretence: we pretend that there is such a person as Hamlet, and thus make it part of the pretence that “Hamlet” is a name for Hamlet (though this is not really so). So far, this allows for “Hamlet exists” to be used only within the pretence; but the role of “really” is to enable it to have a “serious” use as well, for:

“Really” is a word which, when prefixed to a sentence, produces a sentence such that an utterance of it is true (absolutely) if and only if the sentence preceded by “really” is itself such that there is a proposition expressed by it when it is uttered as a move in the relevant game of make-believe, and this proposition is true (absolutely) . . . (p. 370).

In the normal case of telling fictional tales, Evans is clear that propositions are not expressed, though it is make-believedly the case that they are. On a natural reading of the quoted passage (which comes close to the end of a

hastily concluded chapter, which would no doubt have been elaborated considerably had Evans lived), Evans is saying that there is also another kind of case, in which the very act of make-believing to express a proposition is at the same time an act of genuinely expressing a proposition. This interpretation is supported by the fact that most of the discussion attached to this account of “really” relates it to a case of that kind.

Two men . . . both seem to see a little green man on a wall, and are persuaded, with reason, that there is no such thing; they are, they think, victims of a trick . . . [But suppose in fact] there *is* a little green man on the wall. It seems clear that a subject in this situation . . . would actually be thinking of that little green man. (1982: 360–2)

In this case, an utterance of the form “That little green man is *F*” would express a genuine proposition and “‘Really (That little green man is *F*)’ would be true (if the little green man is *F*)” (p. 371). Truth of “Really (*p*)” requires that “*p*” play a double role, expressing a genuine proposition in the course of a make-believe in which in general there is only the pretence of proposition-expression and not its reality.

Dummett (1983) has, with some reason,<sup>11</sup> cast doubt upon whether “That little green man is *F*”, uttered under the circumstances of supposed but non-actual hallucination we have described, is such a sentence, that is, one concerning which it is fictionally the case and also genuinely the case that it expresses a proposition. Even granting the justice of this doubt for this case, there are certainly others for which it would not be just. Perhaps in one of the Holmes stories there occurs the sentence “Baker Street is north of Oxford Street”. This is *really* true (as opposed to so many other sentences in the book): Baker Street really is north of Oxford Street. It would seem that Evans’s account handles this nicely. The narrator of the story, even if he had primarily story-telling intentions, has in fact, whether he cared or not, said something really true. (For it to be the truth we take it to be we would, of course, have to credit the narrator with intending to use “Baker Street” of the real Baker Street, and “Oxford Street” of the real Oxford Street, so bringing these streets into the story, in a way commonplace in fiction.) So a criticism of Evans should not be based merely on the claim that no sentences can live the kind of double life needed to verify “Really (*p*)”.

Dummett has a different objection. Using “E(*g*)” to abbreviate “That little green man exists”, he argues that, given Evans’s account of “really” quoted above,

11 Dummett quotes Evans’s own words: “a *necessary* condition for a speaker to have referred to an object by the use of an expression is that it be the intended referent of that use of the expression” (Dummett 1983: 300, Evans 1982: 318). Cf. also essay II above, p. 69.



the condition for the truth of “Not (really (Eg))” ought to be . . . that either “E(g)” does not express any proposition when used in the make-believe, or it does express such a proposition, but that proposition is not literally true. So understood, however, the negative statement would not genuinely be exploiting the make-believe game; and that was not Evans’s intention. (Dummett 1983: 303)

Dummett’s criticism brings to light an unclarity in Evans’s theory. On one interpretation of his words, “the relevant game of make-believe” is a Russellian description with narrow scope relative to the biconditional. The truth condition, thus construed, is that there is a relevant (perhaps contextually determined) game of make-believe within which “*p*” expresses a true genuine proposition. On this construal, it seems that Dummett is right so say that the condition does not achieve what Evans hoped. A test is that one could put any old rubbish for “*p*” and the condition would not be satisfied, so that “really (*p*)” would be false. The conclusion is that there is no need to understand “*p*” to understand “really (*p*)” which, as Dummett says, was certainly not Evans’s intention.

On an alternative construal, the talk of “the relevant game of make-believe” serves as a presupposition to the truth condition for “really (*p*)”, in the sense that if there is no such game, there is no truth condition. If we put rubbish for “*p*”, no truth condition will be attributed. On this construal, no ground has yet been provided for thinking that Evans’s account fails to do justice to the requirement that some kind of understanding of “*p*” is required for understanding “really (*p*)”. However, Dummett’s criticism can be taken further.

Evans cannot simply require that “*p*” in “really (*p*)” needs to be understood, because if “*p*” contains a fictional name it cannot, on his Russellian theory, be understood. His idea is that, if the presupposition that there is a relevant game of make-believe is satisfied, “*p*” can be quasi-understood, where this is not a species of understanding. But while the presupposition ensures the possibility of quasi-understanding, it does not ensure its actuality. To know that “really (*p*)” has a truth condition you need to know that there is a game of make-believe in which “*p*” has a use; to know what it is for “really (*p*)” to be true you need to know what it is for “*p*”, thus used, to be true. But this cannot in general be known: where “*p*” has no truth condition (as opposed to it merely being the case that F(“*p*” has a truth condition)), there is nothing that counts as knowledge of what it is for “*p*” to be true.

What this criticism brings to light is that Evans has not satisfactorily explained the “game-to-reality shift”. The truth-condition for “really” has to be general, and to shift us in to reality. Yet it remains to be explained how quasi-understanding can be an input to genuine understanding: we cannot know what it would be for “*p*” to be really true in the most common case,

that in which there is no such thing as what it is for “*p*” to be true (absolutely).

We might think that what is needed is a revised account of the truth-conditional impact of “really”. I shall suggest the more radical view that “really”, at least as used in ordinary English, makes no contribution to truth conditions. Everyone can grant that a speaker will normally reach for the word only when he takes it that some fiction, speculation or error is salient and calls for exposure. But it will also be granted, even by those who see “it is true that *p*” as coinciding in truth conditions with “*p*”, that a speaker will reach for the longer form only when, for example, conceding something (“while I admit that it is true that *p*, I would not go so far as to say that . . .”).<sup>12</sup> We therefore need a positive reason to write a role for “really” into truth conditions, as opposed to treating it as a pragmatic indicator. Since it must be factive, so that we have every instance of “if really (*p*) then *p*”, the question resolves into whether we can find a case in which “*p*” is true but “really (*p*)” is false. (We ought to be pessimistic, for it is hard to see how we could have “not really *p*” without “not *p*”.)

Evans suggested such a case: “Had this man’s parents not met, this man would not have existed” contrasts with what results from inserting “really” before “existed”. This example is problematic *ad hominem*, as Evans accepts from Dummett that sentences which coincide in truth conditions may embed differently in modal contexts. Setting this aside, it is not clear that inserting “really” produces any change of truth value, as opposed to some oddity. The interpreter of the “really” version looks for some pretence or speculation or false view to the effect that this man would have existed under the supposed circumstances, and finding none feels dissatisfied; but it would be hard to *dissent* from the conditional supplemented with “really”. So it remains to establish that “really” makes a difference to truth conditions.

“Really” as used in ordinary English certainly does not operate like Evans’s “really”. Rather, as it is natural to expect, it serves to highlight a contrast between reality and any of the things that can differ from reality, which include fiction but are not confined to it. It prepares the interpreter for a contrast with something which, apart from fiction, may be error, or mere appearance, or exaggeration, or understatement, or other departures from reality. Here are some examples.

*Exposing straightforward error:*

The author is quoted as saying he feared a fiend, but what he really said was that he feared a friend.

12 A comparison with “actually” would in some ways be more to the point. But Evans (1979) has quite special views about this word.

*Contrasting the genuine with the feigned:*

The service is currently exploited by a wide range of applicants, but should be confined to those really in need.

*Deflating exaggeration:*

In saying that the number of Jews gassed in Auschwitz was really little over 2 million, as opposed to the frequently claimed 3 million or more, I do not mean to diminish the horror of the events.

*Introducing the seemingly unbelievable:*

something so extraordinary is said that it is as if its denial were already part of the background, and the affirmation implicitly contrasts with it: In 1945, Hitler really expected England to be willing to form an alliance with him against Russia.

*Pre-emptive strike against contrasting views:*

Even idioms it might be tempting to dismiss as mere idioms, not needing to be brought into a unified explanation, can be accounted for: He’s really nice. (Let no one tell you otherwise! And don’t imagine he merely appears to be nice!)

*Contrasting correct and incorrect views about fiction:*

Hamlet is indecisive and morose, but not really neurotic.

The role of “really” in such a sentence is to suggest the concession that the view that Hamlet is neurotic has some followers; if this were not so, there would be some inappropriateness in the affirmation. It does not, however, take one out of the “world of fiction”; it does not effect a game-to-reality shift.

In all these cases, the omission of “really” would appear to make no difference to truth conditions.

Given that it is really rather obvious that the ordinary English “really” does not function at all as Evans’s “really” does, it is better to construe his expression as a technical neologism. The question then is whether it is coherent, and whether it is required. As we have seen, there are doubts about whether it is coherent by Evans’s own standards. From a Fregean perspective (using “Fregean” in the rather special sense mentioned earlier) it is far from obvious that it is required. For Evans, it was supposed to address the problem of how names which couldn’t be genuinely understood could be used in sentences which were genuinely true, notably, negative existentials. The Fregean regards the names as genuinely understandable, so the problem does not arise for him in this form. Frege himself offered a second-level account of existence. For those who prefer a first level account, there is a straightforward option: unlike Frege, who regarded every simple sentence containing an empty name as

lacking truth value (assuming the name to be used as a name, rather than to introduce a concept), treat every atom containing an empty name as false, and deny the unrestricted validity of existential generalization.<sup>13</sup> Once the genuine intelligibility of empty names is granted, both first- and second-level approaches to existence can do without Evans’s neologism.

Whereas Evans assigns the “really” operator wide scope over “Hamlet exists”, Wiggins (1995) assigns it narrow scope. His overall aim is to suggest that a second-level approach to existence sentences deserves development; and his development takes for granted a view of names which has some similarities with Evans’s. His account has the advantage of dealing explicitly not only with fictional sentences, but also those relating to speculation, like Le Verrier’s speculations about Vulcan, speculations in effect dismissed as unsound by a sincere utterance of “Vulcan does not exist”. While I am sympathetic to the idea of developing a second-level approach (and to much else in Wiggins’s paper), I am less persuaded that “really” has any significant role to play. Given what has gone before, I shall assume it is best to see the word as a newly introduced technical device, so the critical questions are not whether it works like its English homonym but whether it is coherent and necessary.

Whereas for Evans, considerations about quasi-understanding made it essential that a fictional name in a true negative existential sentence occur within the scope of “really”, Wiggins allows that quasi-understanding is a kind of understanding. It is made possible only thanks to the fiction, speculation or error, but, Wiggins insists (and I agree) it “can be just as good as understanding ever can be” (p. 107). From the point of view of their intelligibility, empty names can, on Wiggins’s theory, occur anywhere; and Wiggins exploits this possibility by according them wide scope relative to “really”. In particular, he sees “Vulcan does not exist” as containing an occurrence of “Vulcan” which lies outside the scope of “really”, rather as things seem in “Vulcan doesn’t really exist”. We are invited to think of an utterance of the sentence as involving two phases. In the first, we are in “speculative mode” or “rehearsing mode”: we enter into Le Verrier’s speculations, associating our own use of the name with his. In this way, we introduce the concept of being Vulcan. In the second phase, we are in “reality-invoking mode” or “commenting mode”: reality bites, and we deny that this concept is instantiated.

What “really” marks is the transition – within one sentence – from one sort of thinking and talking, the speculative mode, to the other sort, a reality-invoking mode. What the true negative existential sentence will force us to recognize is that we need both these modes . . . . (1995: 106)

13 This “negative free logic” is described by Burge (1974).

In summary form, “Vulcan does not exist” is represented by Wiggins as:

as regards Vulcan, there isn’t really <of> it (Wiggins 1995: 108)

where the “<of>” indicates that we are treating the apparent singular term that follows as introducing an individual concept in the kind of role that makes it a suitable argument for the second level denial of instantiation.

On Wiggins’s account, there are two ways in which a name may be used in the course of saying something true even if it does not have a bearer. One is by being used only in rehearsing mode, in the manner stressed in connection with “as regards Vulcan”. The other flows from the second-level account of existence, which requires that there are occasions on which names, or what appear to be names, or what appear to be pronouns anaphoric on names, in reality function as concept-words, introducing the individual concept corresponding to the name, to which a second-level existence concept is applied. Both these ways are exploited in the account of “Vulcan does not exist” or “as regards Vulcan, there isn’t really <of> it”. The initial occurrence of “Vulcan” does not need a bearer because it occurs in the scope of rehearsing mode. Its second occurrence, or the occurrence of the “it” which depends upon it, does not need a bearer thanks to its interaction with “<of>”, which ensures that it is functioning as a concept-word. This suggests that we should think of “really” and “<of>” as a fused idiom.

For suppose we do not. Then we would have to give an account of sentences of the form “( . . . ) really (—)”, with “<of>” not occurring in (—), which has the consequence that they can be true when a non-denoting name, used as a name, occurs within ( . . . ), but not when such an expression, or a pronoun dependent upon it, occurs within “(—)”; for, by hypothesis, the separable “<of>” alone made possible the non-referring use within “(—)”. These assumptions would make anaphoric dependence between the first and second clauses unintelligible. For example,

Santa Claus does not really bring presents,

with “really” instantiating the technical notion under discussion, would have some such logical form as

As regards Santa Claus, really he doesn’t bring presents.

Here the “he” is not interpretable in accordance with the proposal, for it is in the “reality” part of the sentence and so needs a bearer; yet its anaphoric dependence on the earlier occurrence of “Santa Claus” ensures that it has whatever bearer “Santa Claus” has, that is, none. Because we have supposed that “really” can occur without “<of>”, we cannot appeal to “<of>” to restore an anaphoric connection.

Now suppose that we treat “really” and “<of>” as a single idiom, so that “really” will appear only when we are introducing a name functioning as a concept-word. Then it is unclear that we need anything else from such an expression than that it convert this occurrence of a name into a concept-word-occurrence. If this is done, nothing else is needed, for the name does not need a bearer for the sentence to be true.

So although I am attracted by Wiggins’s proposal, I cannot really see how it works in detail. Moreover, it has a consequence I find strange. Wiggins denies that we can move to such claims as

In reality, Vulcan does not exist

on the grounds that we would then be committed to the intolerable conclusion that really *there is* something to the concept of which nothing answers. One can accept as true

Vulcan is something to the concept of which nothing answers

with the first word genuinely a name (as opposed to concept-word) but used only in rehearsal. Likewise we can accept the existential quantification to

There is something to the concept of which nothing answers

though the initial quantifier is still used only in rehearsing mode, and no more really existentially commits us than would an utterance (as it would most naturally be interpreted) of “There is a detective at 212B Baker Street”. But if the singular sentence were prefixed by “really”, its truth would require “Vulcan” to be used both in commenting mode and as a name (as opposed to concept-word); so its truth would require the existence of Vulcan. Likewise the inferred existential generalization would be in reality mode and not just rehearsing mode.

However, “in reality, Vulcan does not exist” strikes me as true, and I regard this as casting further doubt upon Wiggins’s story. Intuitively, it certainly seems that we can say: here is how reality is: Neptune belongs to it but Vulcan does not. Indeed, even on Wiggins’s own account it would seem that there should be a true reading of the sentence, on which it is equivalent to the claim that it is real that the Vulcan-concept isn’t instantiated. In other words, there seems no reason why the other potential explanation of how an empty name may be used in speaking truly, namely that it functions as a concept-word, should not come to the fore, and do all the work required.

I would like to endorse Wiggins’s view that the proper use of any name, including an empty one, involves immersion in the practice of using that name. In the case of “Vulcan”, the relevant practice is fixed by Le Verrier’s failed speculation, so that a typical user of the name will have some familiarity with that

speculation.<sup>14</sup> What at the moment I cannot see is how this requirement upon the intelligibility of the name can properly be held to be explicitly drawn upon, as opposed to presupposed, in the semantics for sentences in which it occurs. I offer this remark in the hope that Wiggins will supply some more detail about the roles of “really” and “<of>”.<sup>15</sup>

14 This is not necessary, only typical: cf. essay XII below.

15 Thanks to audiences at the University of London Philosophy Programme’s one-day workshop on Reference (February 1998), at Rutgers University’s Philosophy Department seminar, at the 1998 World Congress of Philosophy, and at the Institute of Philosophy at UNAM (The Autonomous National University of Mexico), and the Philosophy Department seminar at the University of Stirling for helpful suggestions; and to Brian Loar, Michael Martin, Stephen Neale, Gabriel Segal, Peter Sullivan, Yannis Stephanou, Jonathan Sutton, Charles Travis and David Wiggins.

Thanks also to the British Academy for contributing to the cost of my travel to the World Congress of Philosophy in Boston in August 1998, where parts of this paper were presented. Wiggins responded to this paper: see Wiggins 1999.

# Knowing meanings and knowing entities

---

## I Acquaintance with meanings

The main claim of this essay is that meanings are not entities; more cautiously, it is the claim that the hypothesis that meanings are entities can make no contribution to an understanding of meaning. This claim is familiar and one might suppose it needed no repetition. However, the idea that meanings are entities still seems to be considered a live option. I hope to make some slightly different attacks upon it, and also to suggest a positive account of how knowledge of meaning is to be understood if meanings are not entities.

I shall assume that meaning and understanding are correlative: meaning is whatever must be accessed in understanding. Consequently, if meanings are entities, they must be essentially involved in an account of understanding. Since understanding an expression involves knowing its meaning, the most straightforward way to connect the claim that meanings are entities with understanding is to construe “knows its meaning” as having the same overall logical form as “loves Mary”: that is, to identify understanding an expression with knowing the entity which is the expression’s meaning.

There is certainly an extensional two place relation of knowing, the relation which Russell had in mind when he developed the notion of acquaintance.<sup>1</sup> Let us use an asterisk after the word to fix this two-place relational use. Knowing\* something normally involves having encountered it, or having had some special kind of epistemic contact with it, the kind of contact that engenders cognitive abilities concerning it. If John knows\* Paris, presumably he has been there and knows his way around; and if Paris is Europe’s smallest capital, John knows\* Europe’s smallest capital, whether or not he knows *that* he knows\* Europe’s smallest capital. The most straightforward account of how meanings as entities figure in understanding is that meanings are objects of this relation: knowing what an expression means is knowing\* the entity it means.

1 Russell (1912b: 23) claims that although “know” is ambiguous in English, the relevant sense is unambiguously expressed by the German by “kennen” and the French “connaître”.



A Wittgensteinian idea might suggest in a general way that there is something wrong: meanings can be stated, but entities cannot be (they can only be named or otherwise referred to). A more detailed argument against this view would draw upon the fact that the extensionality of the knows\*-relation would permit inferences from truths to falsehoods. Since, for example, “chien” and “dog” are synonymous, the entity view must say that each means the same entity:

(1) The meaning of “dog” = the meaning of “chien”.

For Pierre, a monolingual French speaker, to understand “chien” is (on the entity theory) for him to know\* the meaning of “chien”:

(2) Pierre knows\* the meaning of “chien”.

These ought to entail

(3) Pierre knows\* the meaning of “dog”.

Given that knowledge\* is being used to explicate knowledge of meaning, (3) amounts to the claim that Pierre knows the meaning of “dog”; but since he is monolingual, this is false. So either (1) or (2) is false. So either meanings are not entities, or else they are entities which connect with understanding in some way yet to be made clear.

An objection is that this argument would prove too much. Compare it with:

(4) The reference of “Hesperus” = the reference of “Phosphorus”.

(5) John knows the reference of “Hesperus”.

(6) John knows the reference of “Phosphorus”.

In this case, the objection runs, (4) is unquestionably true; but (5) could be true without (6) being true. John might never have encountered the word “Phosphorus”; or might have done so without realizing that its reference was the same as that of “Hesperus”. However, it would be folly to conclude that Venus (that is, the reference of “Hesperus”) is not an entity.<sup>2</sup> Responding to this objection involves distinguishing ways in which what may appear to be cases of knowledge\* are not really so; and distinguishing reference from referent.

Many constructions of the form “X knows the F” do not express knowledge\*. For example, if the police do not know *the identity of the murderer*, we

2 Thanks to David Chalmers for this objection.

should not think of this as an example of failure of knowledge\* (i.e. an example of ignorance\*), but as something expressible only by a non-extensional idiom. What the police do not know is who the murderer is. In such an indirect question, the non-extensionality of “knows” is active, for it may be that Sam Jones is the murderer and the police both know who he is (they know his identity) and also know\* him, but little suspect that he is the murderer. So ignorance or knowledge of identity is not ignorance\* or knowledge\*. We can hear a sentence like “John knows the capital of France” as equivalent to “John knows which (city) is the capital of France”. Thus understood, the sentence does not express knowledge\*, and it may be that Paris is the smallest capital of any country in Europe, that John knows that Paris is capital of France and yet that John does not know that the smallest capital of any country in Europe is capital of France. So the first point is that if knowing the reference of “Hesperus” is heard as equivalent to knowing what it is that “Hesperus” refers to, the argument in (4)–(6) does not mirror that in (1)–(3).

We are helped to hear (4)–(6) as involving a notion other than knowledge\* by the use of “reference” as opposed to “referent”. The referent of an expression is the object itself; by contrast, the reference of an expression is more fact-like, and so is apt to suggest *knowledge that* rather than knowledge\*. One can *state* the reference of “Hesperus” (e.g. by saying that it refers to Hesperus); but Hesperus itself, that is, the object to which “Hesperus” refers, that is, the referent of “Hesperus”, is not the sort of thing that could be stated. If we replace “reference” by “referent” in (4)–(6), and keep to knowledge\* throughout, then the argument is, I believe, valid. It is just as valid as a corresponding argument in which “sees” replaces “knows”: that is the sort of relation knowledge\* is supposed to be. So the objection that (1)–(3) proves too much fails.

Expressions like “knows the meaning of ‘dog’” do not express knowledge\*. Consequently, it would be wrong to regard “the meaning of ‘dog’” as a referring expression or as a denoting phrase: either construal would liken the overall logical form of “knows the meaning of ‘dog’” to “loves Mary”, and this is what sets up the premises for the defective inference of (1)–(3).

The argument in (1)–(3) applies straightforwardly to Fregean senses.<sup>3</sup> If we take seriously Frege’s view that senses are entities, then grasping the sense of an expression, which is supposedly necessary and sufficient for understanding it, would be a relational matter: “grasps” would induce the same logical form as “knows\*”. So one could replace “knows\*” by “grasps” in (1)–(3) and arrive at the unpalatable conclusion: the sense of “dog” = the sense of “chien”, but one may grasp the sense of “chien” without grasping the sense of “dog”.

Philosophers of our own time who have been concerned to develop Frege’s

3 This paragraph and the next were prompted by a question from Ed Zalta.

views (for example Dummett, McDowell, Evans) have seen grasping a sense as knowing a fact, for example, the fact that something satisfies “dog” iff it is a dog. In these facts there is no reference to such entities as senses.<sup>4</sup> The view that knowledge of meaning is knowledge of fact is not the only alternative to the view that knowledge of meaning is knowledge\* of an entity. It might be that the knowledge is practical, not necessarily formulable in terms of a fact or proposition known. Both these alternatives to knowledge of meaning as knowledge\* of entities are left open in what follows.

## 2 *x* means *y*

The claims so far establish only that either meanings are not entities, or else they are entities which connect with understanding in some way yet to be spelled out. In order to be better placed to consider alternative ways of incorporating meanings as entities into an account of understanding, I will introduce another two-place extensional relation, that of meaning, holding, supposedly, between expressions and the entities they mean. To signal this use, let us call the relation “meaning\*”. Horwich (1998) argues that meaning\* is (a) a relation which occurs outside of semantics, and (b) is the right relation for connecting expressions and the entities they mean. Yet he agrees, essentially for the reasons I have offered, that even though “dog” means\* DOG (the capitalization is introduced by stipulation to refer to the entity which is the meaning of the corresponding lower case expression), knowing that “dog” means DOG does not suffice for understanding “dog”. Why, then, bother with meanings as entities?

Concerning the “use” theory of meaning which Horwich aims to establish, he says that it is a point in its favour that it accommodates our ordinary way of speaking of *meanings* as a species of entity to which words stand in the relation “*x* means *y*”. Moreover, it makes do with the familiar, non-semantic use of the word “means”. When we say, for example, that black clouds mean it will rain, or that the expression on his face means that he is sad, we are deploying a notion of *means* which is, roughly speaking, the notion of *indication*. To say, in this sense, that *x* means *y*, is to say, roughly, that *x* provides a good reason to believe in the presence of *y* (Horwich 1998: 47).

4 Many of Frege’s main claims can be phrased with recourse merely to the same-sense relation, without invoking senses as entities. An exception is his view of indirect reference, if customary senses as entities are not available as *Bedeutungen* in these contexts. But a highly Fregean position on indirect discourse can be developed just in terms of substitution conditions, without senses as entities. If there were a single relation of synonymy, and if it were an equivalence relation, there could be no more objection to meanings as entities than to directions as entities (whatever objections there might be to using meanings in an account of understanding). I do not accept the antecedents of the conditional, though that is a topic for another occasion.

Though the symbolism “ $x$  means  $y$ ” is presumably supposed to introduce meaning\*, with the variables marking positions fit to be filled by singular terms, Horwich’s examples strike me as inappropriate. “Black clouds mean it will rain” has “it will rain” in  $y$ -position, and the subsequent example has “that he is sad”. Although there are views according to which sentences mean entities, or do so when prefaced by “that”, these are highly controversial, and cannot be used for his purpose of showing that quite ordinary beliefs sustain the view that meanings are entities.

Horwich’s gloss on what, I presume, is the means\*-relation also fails to apply to the examples. It is not grammatical to affirm either that black clouds provide a good reason to believe in the presence of it will rain, or that the expression on his face provides a good reason to believe in the presence of that he is sad. The English “means” can be used for purposes other than to express a two-place relation between entities. So we remain in need of a reason for thinking of meanings as entities.

### 3 “That”-clauses do not refer to meanings

Horwich (1998) suggests that whole sentences stand in the means\*-relation to their meanings, and that this fact is exploited in propositional attitude ascription. On this view, if I say that John believes that snow is white, “that snow is white” means\* that snow is white, and John’s belief consists in a relation to this meaning-entity. This gives the basis for constructing the following argument whose conclusion evidently could not be true, though someone of Horwich’s persuasion must accept that the premises could be:

- (7) The meaning of “snow is white” = that snow is white.
- (8) Pierre is amazed that snow is white.
- (9) Pierre is amazed the meaning of “snow is white”.

The argument refutes only a conjunction: meanings are entities and are referred to by the “that”-clauses of propositional attitude ascriptions.<sup>5</sup>

5 This kind of argument is used by Moltmann (forthcoming) against the view that propositional attitudes are relations to propositions: “Pierre is amazed the proposition that snow is white” does not follow from (8) together with the assumption that, in (8), “that snow is white” refers to the proposition that snow is white. Paul Horwich pointed out that the same considerations would tell against various other accounts of propositional attitude ascription, including Davidson’s. To show that there is a difficulty for the paratactic analysis, one needs no identity premise: it is enough to observe that whereas “John believes that” or “John said that” are, in a suitable context, capable of making an assertion (with the demonstrative referring forwards or backwards), the same does not hold for “Pierre was amazed that”. This suggests that it is worth exploring the view that there is no unified category of “propositional attitude ascriptions”.

## 4 Meanings and understanding

Returning now to ways of using meanings as entities in an account of understanding, an obvious defect in the simple way envisaged in section 1 is that knowledge\* of meanings as entities doesn't connect an expression to the known\* entity. A better theory might make the following claim:

- (10) Someone knows the meaning of an expression, *e*, iff the person knows that there is an entity *x* such that *e* means\* *x*.

This moves away from a two-place extensional relation of knowledge\* to the more familiar non-extensional *knowledge that*, but retains the entity account of meaning by invoking meaning\*. I do not argue that this version of the entity theory leads to straightforward contradictions, as the first does; only that entities turn out to play no essential role in a viable version of this theory.

(10) cannot be what we are looking for, as it expresses no more than knowledge that the expression means something, as opposed to knowledge of what it means. A serious candidate must have a more *de re* character, perhaps the following:

- (11) Someone knows the meaning of an expression, *e*, iff the person knows, concerning some entity *x*, that *e* means\* *x*.

A standard view is that although substituting co-referring names in a “knows that” context does not guarantee that truth will be preserved, names in such contexts do provide a basis for existential generalization. On this view, “Pierre knows that Marie is engaged” entails “There is someone who Pierre knows is engaged”. Now suppose that the capitalized expressions used by Horwich to refer to meanings are names. Then

- (12) Pierre knows that “dog” means\* DOG

would entail

- (13) Concerning some entity, *x*, Pierre knows that “dog” means *x*.

Suppose Pierre is our monolingual Frenchman, and that (12) is true of him because he knows how to operate the capitalization convention, and he does, in fact, know\* the meaning of “dog”, thanks to knowing\* the meaning of “chien”. If the entailment holds, then (13) would be true and would provide a counterexample to (11): there would be something such that Pierre knows “dog” means it, even though he does not understand “dog”, that is, does not know its meaning. However, this argument raises some difficult questions.

Arguably, “DOG” functions like a definite description, in a way that does not support existential generalization across “knows that”. If Pierre knows that Marie is engaged to the fiancé of Marie, there is room for doubt whether it follows that there is someone whom Pierre knows Marie to be engaged to. This latter, *de re*, ascription of knowledge requires the knower to know\* the relevant object, which in turn entails having a good range of information about the object. This explains why the inference is arguably valid when the position of the existential variable in the conclusion is filled by a name in the premise: understanding a name involves associating it with a fairly rich body of information. But in the case of a description, like “the fiancé of Marie”, one may understand the expression without associating any further information with it; one may understand the description without knowing\* its denotation. In such a case, it would be true that Pierre knows that Marie is engaged to the fiancé of Marie, yet false that there is someone of whom Pierre knows that Marie is engaged to him. So if “DOG” functions like a description, the inference from (12) to (13) is invalid, and we no longer have a counterexample to (11).

As we envisaged Pierre, in the discussion relating to (12) and (13), the entity view must allow that he did in fact know\* the meaning of “dog”, thanks to understanding “chien”. So the relevant question is not just about existential generalization into “knows that” contexts, but rather concerns the following inference pattern:

- (14) S knows\*  $x$ .  
       S knows that the  $G$  is  $F$ .  
       The  $G = x$ .  
 So: For some  $x$ , S knows that  $x$  is  $F$ .

An example not connected with meaning: suppose that Pierre knows\* Marie’s fiancé: he is a colleague at work, though if asked in the presence of this colleague whether that man was due to marry Marie, Pierre would profess ignorance. He does, however, know that Marie is engaged to the fiancé of Marie (having inferred it soundly from his knowledge that Marie is engaged). Is there someone who Pierre knows is engaged to Marie? It is tempting to answer affirmatively, for the idea was that knows\* is an extensional relation, and that knowing\* an entity will justify a *de re* ascription of knowledge concerning it. If this is not right, then at a minimum the entity theorist must offer some serious further theorizing on the nature of *de re* ascriptions of knowledge.

If, however, the suggestion is right, so that the pattern of (14) is valid, then the original counterexample stands. Pierre knows\* the meaning of “dog” (thanks to knowing\* the meaning of “chien”), and knows that the meaning of “dog” is the meaning of “dog”. So there is something such that Pierre knows that it is the meaning of “dog”. The right hand side of the relevant

instance of (11) is satisfied, but not the left, for Pierre knows no English. This does not show there is anything amiss with the supposition that meanings are entities; only that such entities have not yet been found a role in an account of understanding.

## 5 Indirect questions and knowing what

The view that meanings are entities may be encouraged by theory, for example, by the view that meanings as entities help one explain understanding, or that the connection between language and the world requires the meanings of words to be entities in (typically extra-linguistic) reality. The first of these I have attacked; the second, though I believe it to be mistaken, cannot be addressed here. The entity view is also encouraged by certain idioms. The seemingly ubiquitous “means” (in the sense of means\*) has already been discussed. In this section, I will mention two other idioms, and also point to the form (though not the content) of an entity-free account of knowledge of meaning.

Meanings, it seems, can be counted, and questions of identity and difference arise. For example, an ambiguous expression is one with more than one meaning, and we can wonder if the meaning of “chien” is the same as the meaning of “dog”. This, I suggest, is mere idiom, which should not drive us to entities. There are relations between expressions of being alike or different in meaning, but we are not obliged to think of the respects of similarity or difference as further entities. “Bank” applied to financial institutions does not mean the same as “bank” applied to sides of rivers; but we need not analyse this as saying that financial “bank” is related to a different meaning-entity from that to which river “bank” is related. Similarly, the fact that to do it for John’s sake is not the same as to do it for Mary’s sake does not require us to think of sakes as entities.<sup>6</sup>

The idioms most apt to encourage the entity view have the form “what *e* means” or “the meaning of *e*” as in “John knows what ‘dog’ means” or “John knows the meaning of ‘dog’”. There is a temptation to construe these as referring (or at least as denoting) expressions, satisfied by entities. As we have already seen in other cases (knowing the identity of a person), such appearances often mislead. In the present case, I suggest that “the meaning of *e*” can always be rephrased as “what *e* means” and that this latter is an indirect question (whose direct form is “what does *e* mean?”). It is not a referring or denoting expression, and nor is a referring or denoting expression normally usable in answering the question. Once again, the Wittgensteinian idea provides the explanation: meanings can be stated, but entities cannot be.

6 I believe that synonymy can be used to provide a good understanding of such claims as that “dog” might have meant something different, without appeal to meanings as entities.

*Knowing what* in some ways resembles *knowing who*. In both cases, a sufficient condition for possessing the relevant knowledge is knowing a satisfactory answer to the questions which these idioms indirectly formulate. In the case of *knowing who*, contextual factors play a large part in determining what counts as satisfactory. Often, a proper name of the right person counts as a good answer; but we can also fail to know who N is (where “N” holds the place for a proper name).

In the case of knowing what an expression means, there is less room for contextual variation, but also less clarity about what would constitute a satisfactory answer. Indeed, to insist upon an answer formulable in words seems mistaken. Infant language users may show that they understand what an expression means, without being at all good at answering a question of such a sophisticated kind.

I suggest we take our cue from cases like this: he knows what to do in an emergency. In an emergency, the question may arise: what is to be done? Someone who knows what to do may not have formulated the question in words, and may not be able to formulate a correct answer in words. Perhaps one has to tie a bowline, but he might be incapable of describing the complex manual process this involves, especially if the situation is stressful. However, if he does indeed know what to do, his behaviour will demonstrate an answer. Someone who can reliably (non-accidentally) demonstrate an answer has the knowledge in question. Similarly in the case of language: knowing what an expression means is being able reliably to demonstrate in one’s behaviour a correct answer to the question “what does it mean”?

One may know that such an answer has been demonstrated without oneself knowing the answer. The boat is righted, the emergency is over. Evidently what he did saved the day: he showed he knew what to do. But we may know all this without knowing exactly what he did (let alone that it involved tying a bowline). One may know that someone has demonstrated a correct answer to a “what does it mean” question without knowing an answer oneself, as one can be sure that the foreigner whom one does not oneself understand does understand his own language. Demonstration of an answer will only convey the answer to adequately equipped observers, just as pointing out someone in a line-up will only convey to adequately informed observers that that man is the murderer.

It is another issue whether or not it is possible to say anything in more detail, yet still at a level of generality that would hold for every linguistic expression, about what demonstrating a correct answer consists in. A large variety of approaches is left open, including Horwich’s use theory. The proposed shape of knowledge of meaning is neutral even on such questions as whether the capacity to demonstrate an answer to questions of the form “what does *e* mean?” can be expressed (or in some way captured) in terms of propositional knowledge, explicit or implicit. The point of the present remarks is not to try to take these issues further, but only to show that certain



idioms (like “knows the meaning”) which might seem to call for entities as meanings do no such thing.

## 6 Things which are not entities?

Meanings are not entities, but a huge number of expressions mean something, and, typically, different expressions mean different things. How are we to understand these last apparent quantifiers (“something”, “different things”)? Many metaphysicists have addressed the fully general form of this question (often in rather different terminology). Two contrasting views refuse to allow a distinction between things and entities. A lean, Quinean, version is that everything is an entity and so, given that meanings are not entities, there are no things which expressions mean: we must say instead something like “expressions are meaningful”. A more exotic version is that, precisely because everything is an entity, we must count things as entities too, and so, to the extent that we hold that expressions mean something, we must allow they mean entities.

The safest position for me to take, and the one I would insist upon if pushed, is that one cannot systematically distinguish between things and entities. Hence, in rejecting meanings as entities, one must also reject any literal affirmation to the effect that expressions mean some *thing*.<sup>7</sup> However, it seems to me that our ordinary thought does make a distinction, and in a way that treats things as somehow ontologically non-serious. Perhaps this distinction would ultimately have to be rejected, if our metaphysical project were transcendent: the attempt to say what there is, absolutely. But if we are aiming only at immanent metaphysics, an account of the commitment of our actual beliefs, there is at least a *prima facie* case for recognizing things which are not entities. In the following pairs, the first member, using “thing”, states an uncontroversial truth (except possibly (17a), though I believe even this would sound uncontroversial to ears untainted by philosophy), whereas the second is doubtful (in point of truth or in point of intelligibility):

- (15a) Some things simply are not done (e.g. swearing in front of the children).
- (15b) Some entities simply are not done.
- (16a) Some things are quite impossible (e.g. putting a girdle round about the earth in forty minutes).
- (16b) Some entities are quite impossible.
- (17a) Some things don't exist (e.g. Vulcan, dragons).
- (17b) Some entities don't exist.

7 Horwich said that he would not mind whether meanings are called entities or things, so long as one recognized their existence. Within the framework of the present argument, I would insist upon what I have called the safest position.

- (18a) He said some amazing things (e.g. that he is Lazarus, come from the dead).
- (18b) He said some amazing entities.

The methodology would require further refinement, for even those firmly committed to the existence of events and who found “Something unpleasant happened this morning” an impeccable truth might look askance at “Some unpleasant entity happened this morning”. I do not pretend that examples (15)–(18) establish any more than that we have here something (a thing?) worth exploring.

Candidates for things which are not entities, other than meanings, are propositions, truth conditions, sayings, questions and facts. Ways may be more like things than like entities: there may be one or more ways of doing something, but not one or more entities of doing it. This would have consequences for possible worlds (ways the world could have been), qualia (ways of experiencing the world) and properties (ways of classifying the world).<sup>8</sup>

8 This is an expanded version of the first part of my contribution to the 22nd Wittgenstein Conference at Kirchberg in August 1999, under the title “Meaning, Composition, Homophony”. For helpful comments I would like to thank audiences at that Conference and at the first Polish–English Graduate Seminar in Torun (September 1999); and, especially, Paul Horwich.

# Two ways to smoke a cigarette

---

## I Introduction

One of Frege's legacies is his insistence upon the compositional character of language. In his system, the senses of complex expressions are composed of the senses of their parts; and the referent of a complex is determined by the referents of those of its parts which are supposed to have referents. Many philosophers who would not think of themselves as Fregean regard these theses or their analogues as virtually truistic. Certainly, there is a truism in the area: we understand sentences thanks to understanding the words of which they are composed and how these words are arranged. But this is far from yielding anything deserving to be called a "principle of compositionality". Many such principles, as in Frege's own work, take it for granted that meanings are entities, and this is not truistic (as shown in essay X above).<sup>1</sup> All versions assume that there is some correct interpretation of the view that simple expressions have meaning "in isolation", independently of their occurrence in complexes; this seems to be inconsistent with Frege's "context principle": "never to ask for the meaning of a word in isolation, but only in the context of a proposition" (Frege 1884: x). When we reflect upon what animated this remark of Frege's, we become aware of various possibilities of circularity: perhaps the meaning or sense of a simple expression is, by definition, just whatever it contributes to the meaning or sense of the larger wholes into which it enters, in which case some sort of compositionality is true by definition. This may also happen in another way: for the thesis to have any bite, it must be relativized to a syntax for the language, which will tell us when we have an occurrence of a simple expression (there is no occurrence of the word "gnat" in "indignation"). But if a constraint

1 For example: "the meanings of complex expressions . . . are constructed from the meanings of the less complex expressions . . . that are their constituents" (Fodor and Lepore 2001: 350–1). The commitment to meanings as entities is also present in versions which make use of functionality, for example "The meaning of a compound expression is a function of the meaning of its parts" (Janssen 1997: 419). Presumably this requires there to be functions whose arguments and values are meanings.

on syntax is that it identify simple expressions in such a way as to respect compositionality, another kind of circularity looms. The theme of this essay is that the compositionality of natural language is not to be taken for granted, and that an appropriate argumentative strategy is not some attempt at a global proof that our language is or is not compositional, but a detailed study of examples of the apparent failure of compositionality.

Traditionally the main explanandum for compositionality has been linguistic novelty or creativity, something alluded to by Frege in these words:

It is astonishing what language can do. With a few syllables it can express an incalculable number of thoughts, so that even a thought grasped by a terrestrial being for the very first time can be put into a form of words which will be understood by someone to whom the thought is entirely new. This would be impossible, were we not able to distinguish parts in the thoughts corresponding to the parts of a sentence, so that the structure of the sentence serves as an image of the structure of the thoughts.  
(Frege 1923: 390)

Our capacity to deal with novelty was also mentioned by Wittgenstein in the *Tractatus*:

4.027 It belongs to the essence of a proposition that it should be able to communicate a *new* sense to us.

4.03 A proposition must use old expressions to communicate a new sense.

This much is delivered by the truism.

Indexicality shows that novelty in this context is a slippery notion: an utterance may put something familiar to an unfamiliar use. If compositionality is well motivated by our capacity to handle novelty, it should be well motivated by the novel use of familiar words. In short, compositionality ought to apply to expression tokens, and not just to expression types.

## **2 Understanding utterances: covert indexicality as a resource for compositionalist**

One understands utterances thanks to understanding the words used in making them; but indexicality shows that other cognitive resources are involved as well. For example, one must use general cognitive skills to identify the referent of a demonstrative use of a pronoun like “that”. If understanding and meaning are correlative, the need for non-semantic skills to arrive at understanding might suggest a failure of compositionality, for it would be correlative with finding a meaning for the whole which goes beyond the meaning of the parts.

The obvious defence of compositionality in the face of this point will

adopt the kind of structure found in Kaplan's distinction between character and content (Kaplan 1977). Whenever used, the familiar character of an indexical word calls for the exercise of a recognitional skill on that occasion. A skill exercised at the behest of a semantic rule can be classified as belonging to understanding, so the occasion-specific identification of content also belongs to understanding. Very roughly, the rule for "I" is that it is to refer to the speaker of the utterance in which it occurs, and for "elle" that it is to refer to something of feminine gender made suitably salient in the context. To understand an utterance containing a demonstratively used pronoun involves knowing the relevant general rule *and* applying it in the specific context. This conception of understanding seems to compose: one understands an utterance of "Elle a faim" if one uses the rule for "elle" to come to know to whom or what it refers on the occasion, and appreciates that that entity has been said to be hungry. I shall assume that indexicality poses no threat to an appropriately formulated thesis of compositionality, and hence that a defender of compositionality should be happy to classify an apparent counterexample as nothing more than an example of indexicality.

A serious utterance of "It's raining" is normally (perhaps always) taken to say that it's raining at some particular place and some particular time. The temporal element can be classified with "I" and "elle" as a manifestation of overt indexicality: the rule for the present tense says (ignoring such complications as the historic present) that the relevant time is the time of utterance, so understanding the utterance involves understanding it as saying, concerning the time of utterance, that it is raining *then*. But there seems to be no expression which in this way introduces the place of which rain is predicated, so utterances like this appear to be counterexamples to compositionality: as uttered at different places, the sentence has different meanings, but these meanings seem to be a function of the place of utterance and not of the meanings of the parts (cf. Crimmins and Perry 1989).<sup>2</sup>

Although this phenomenon can be treated in more than one way, the response I wish to highlight for the subsequent discussion is that of attributing *covert indexicality*. On this view, at some level of description of the sentence (its "logical form"), there is an implicit variable for places, and understanding an utterance of the sentence requires the understander to identify the appropriate place (cf. Stanley 2000). If indexicality in general is consistent with compositionality for utterances, then so is covert indexicality.

By contrast, the distinction between referential and attributive uses of definite descriptions as a pragmatically determined difference in truth conditions

2 "Our semantics is not compositional, but there is system in the noncompositional mayhem" (Crimmins and Perry 1989: 711). Systematicity is their substitute for compositionality. In fact, on their view it would seem that *propositions* are compositional, even if sentences are not.

arguably does pose a threat to compositionality. The idea is that for some description sentences there is no lexical or structural ambiguity, but proper understanding of an utterance of one involves selecting one or other truth condition, the singular truth condition appropriate to the referential reading, or the existentially general truth condition appropriate to the attributive, and that this is to be done on a holistic contextual basis.<sup>3</sup> This selection is not based upon an understanding of any part of the utterance but is required for understanding. So understanding the whole goes beyond an understanding of the parts.

The compositionalist, therefore, cannot accept every form of “pragmatic ambiguity” as consistent with his thesis, but can accept the form of pragmatic ambiguity induced by indexicality. In what follows, I consider covert indexicality as a way in which a defender of compositionality might deal with apparent counterexamples; though the main thrust of the essay is that most of the counterexamples are not really ambiguous at all. If this is right, the compositionalist has no case to answer.

### 3 Apparent counterexamples to compositionality

In the following examples, the compounds are supposed to be ambiguous, while their parts and manner of construction have no relevant ambiguity. The conclusion is supposed to be that their meaning (any one of the various meanings they ambiguously possess) goes beyond the meaning of the parts, so that they are counterexamples to compositionality. (The examples are consistent with the truism, since even in such cases understanding the words plays some part in understanding the compound, enough to justify “thanks to”.)

- 1 adjectival modification:
  - (a) “Italian book”: book in Italian, book about Italy, book made in Italy, book in the pile to take to Italy?<sup>4</sup>
  - (b) “feline care”: care for, by, of or in the manner of felines?
- 2 genitives:
  - (a) “John’s leg”: one (which?) of the limbs composing John’s body, or the leg which John, a student of anatomy, is dissecting?

3 Presentations of the referential/attributive distinction as this kind of pragmatic ambiguity are given by, for example, Stalnaker (1970: 41–3), Récanati (1993: 282–3), Bezuidenhout (1997). Récanati makes a good case for this having been Donnellan’s original intention, though one misunderstood by many commentators. A compositionalist who accepts Donnellan’s data about the truth values of utterances containing definite descriptions will see this as a manifestation of lexical ambiguity: cf. Evans (1982: 325).

4 Perhaps “Italian” is both noun (denoting a language or a native of Italy) and adjective. These examples are supposed to keep to the adjectival form, though ambiguity about which form is at issue can arise.

- (b) “John’s table”: the/a table John owns, has sold, has bought, has borrowed, has lent, wants to use, often uses, made, . . . ?
- 3 compounded nouns:
- (a) “sand cleaner”: something which cleans sand or which uses sand to clean?
  - (b) “hi-tec management”: management of a business producing or servicing hi-tec things, or management by methods involving hi-tec?
- 4 noun + verb
- (a) “John runs”: with his legs? Or like butter runs when hot, or paint runs when wet, or like the Thames runs from Oxford to London?
  - (b) “John smokes”: cigarettes? or emits smoke (like Etna)?
- 5 The Travis effect:
- (a) “Those leaves are green”, when brown leaves are painted green, may mean something false (if we are interested in natural colour as a sign of something) or something true (if we are merely interested in chromatic matters).
  - (b) “The squash ball is round” may mean something true (as round is its normal shape) or something false (as it is currently in violent contact with the wall and so nearly hemispherical).

In each case, there are supposed to be two or more “readings” or “disambiguations” corresponding to different meanings. But the parts are not relevantly ambiguous, and there is no relevant ambiguity in how they are put together. Hence the various meanings are not determined by the meaning of the parts and how they are combined. Hence a semantic theory could not derive suitable properties for the compounds from suitable properties for their parts.

Taking it for granted that the anti-compositionalist objector is right to say that there is no relevant lexical or structural ambiguity, a defender of compositionality in the face of the counterexamples has various options, and I shall consider just two of them.

- 6 Covert indexicality. Some alleged counterexamples display not ambiguity but different assignments to hidden indexical elements.
- 7 Unspecific meaning. The counterexamples confuse distinct “readings”, that is, distinct meanings, with distinct ways in which one and the same meaning could be true.<sup>5</sup>

The drift of this essay is that option 7 has been given insufficient attention,

5 This option has been applied to alleged ambiguities in sentences with more than one quantifier by, among others, Kempson and Cormack (1981). If their word “interpretation” can be read as “way of being true”, their §6 provides a well-argued case for the kind of position I am applying to the different cases discussed here.

and that it is a good option to take with respect to most of the examples under consideration here. The main argument for favouring 7 is that alternative approaches along the lines of 6 require the determination of a value for the covert variable if understanding is to be possible, and this does not seem to accord with the facts. This establishes at best that 7 is sometimes the best strategy for the compositionalist, but it does not speak to other cases, and does not address the question whether we are somehow committed to finding some successful compositionalist strategy.

## 4 Applying the strategies

### 4.1 Adjectival modification

The objector seems to me right to present cases like 1a and b as ones in which it is very implausible to attribute either lexical or structural ambiguity. First, there are apparently endless ways in which something can count as an Italian book, not all of which could be foreseen in advance in learning “Italian” or “book” or the construction of adjectival modification. If the word is learnt in a context in which the learner appreciates that the contextually salient way for something to be Italian is to be made in Italy, then an ambiguity account would suggest that this is the meaning learnt. It would then be hard to explain the ease with which people go on to use “Italian” in contexts in which other ways of being Italian are salient. Second, if these cases were to be explained by lexical or structural ambiguity, the ambiguity should be available in cases in which it plainly is not. For example, if there were some fixed meaning of the construction which relates what “book” refers to with an intended trip to the place referred to by the noun (“Italy”) which corresponds to the adjective, then “generous book” should have a reading (although perhaps one hard to access in normal contexts) on which it refers to a book in the pile to take to generosity. This is not a hard to access reading, but an absurdity.

If these cases are genuinely ambiguous, the ambiguity is pragmatic, as opposed to lexical or structural. In most contexts hearers will identify the intended relation between “Italian” and “book”; the identification will draw on context, and will exploit considerations of plausibility and relevance. In the present discussion, the question is whether pragmatic ambiguity arising from covert indexicality can provide an adequate account.

On the covert indexical theory, the logical form of “Italian book” could be represented by something like “Italian *R* book”, where the interpretation of the relation variable *R* is to be supplied by the context. The form “*xRy*” could be interpreted so as to be true of the satisfiers of *y* which are written in a language which satisfies *x*, or so as to be true of the satisfiers of *y* manufactured in a place which satisfies the noun from which *x* is formed, and so on. The account has the advantage over accounts in terms of lexical or



structural ambiguity of not requiring that the various readings of the construction be settled in advance. Moreover, context can point to sensible interpretations of *R*, and can place obstacles in the way of accessing interpretations which, in other contexts, would be natural. However, unless nothing better can be found, it seems an extraordinary account as applied to this kind of case. It implies that you would not have understood an utterance like “Let’s read an Italian book together” unless you had identified such an *R*; whereas in fact it seems you do understand even when you are in doubt about *R*. You may go on to ask “Do you mean a book in Italian or a book about Italy?”, but this no more shows that you did not understand the first remark than if, in response to “Let’s go to the movies” you say “Do you mean let’s go tonight or later?”. In both cases, the proposal was fully intelligible but not fully specific.

Another difficulty for the covert indexical account is that once context has made one determination of the variable salient, it should remain salient throughout the sentence unless there are contrary indications, and a shift would produce a special zeugmatic effect. For example, it would be odd to say

8 It’s foggy; cold too

in the hope that context would determine different places for fog and cold, making 8 in effect equivalent to:

9 It’s foggy – here in London; cold too – there in Iceland.

However, there’s at least no firm intuition that there is something odd about the following shift:

10 This is an Italian book – it’s by Ginzburg; and this is another Italian one – it’s in the Lonely Planet series and has lots of information about Italy.

An alternative to the covert indexicality account is to say that “Italian” functions just as one would expect from a dictionary entry: “Italian *F*” is satisfied by a satisfier of *F* of or pertaining to Italy. This is an unspecific but definite, unambiguous and complete meaning. An Italian book is one of or pertaining to Italy, and a book may pertain to Italy by being on my pile of books to take there, or by being about Italy, or by being manufactured in Italy, or by being written in the dominant language of Italy. The different so-called readings of a sentence like “Let’s read an Italian book” do not correspond to different senses or meanings, but to different ways in which “we read an Italian book” can be made true: by reading a book written in Italian or a book about Italy or a book from my pile of books to take to Italy. It is no more correct to

regard this as ambiguity than to regard “John runs” as ambiguous on the grounds that it could be made true by John running in bare feet or in trainers, quickly or slowly, east or west.

It seems to me hard to deny (a) that we should distinguish different “readings” of a sentence from different ways in which it could be made true and (b) that this distinction has not always been scrupulously adhered to. There is plenty of scope for detailed disagreement about precisely how the distinction divides the cases. In order to promote the claims of unspecific but determinate and unambiguous meanings, I will show, quickly and rather dogmatically, how the idea applies to various other cases.

“Feline care” is satisfied, in the same sense, though not in the same way, by a vet tending a cat and by a cat tending the puppy to which she is acting as foster mother. Both vet and cat supply care of or pertaining to felines, the vet by supplying care for a feline, the cat by being a feline which supplies care.

## 4.2 *Genitives*

I claim these are similar: their meanings are highly unspecific, and cover a range of cases. John’s leg or table is a leg or table of or pertaining to John, which a leg or table may be in various ways. There are many different ways in which “John’s leg is on his table” may be true, but the different ways are ways in which the same meaning can be true.

Here is an argument for an alternative account in terms of covert indexicality:

Understanding a singular noun phrase like “John’s table” involves identifying the object to which it refers, but one cannot do this until one knows what relation is supposed to link John to a table. The alleged unspecific meaning is not something that can be understood. On the covert indexicality theory, however, a full understanding requires that the relevant relation be identified, and this will lead to the identification of the relevant object, and without this identification the utterance cannot be understood.

The first premise is questionable. Perhaps one identifies something only by connecting it with antecedently possessed singular information relating to it. In that case, identification is not a prerequisite for understanding: definite descriptions can be understood even if they fail to denote, or if they do denote but the denoted object is unfamiliar to the understander. If the relevant notion of identification is weakened, perhaps to “one identifies an object, as the supposed referent of a noun phrase, only if one can give a satisfactory answer to the question ‘to which object was the speaker, by using that phrase, purporting to refer?’” then it is unclear that a non-parrot-like

repetition of “John’s table” is an unsatisfactory answer. If the standards of what is satisfactory are raised, genuine cases of understanding will be misclassified as cases in which there is no understanding.

A further consideration favouring unspecific meaning over covert indexicality is that one can mix different possessive relations without incongruity:

- 11 This is indeed my book [sc. the one I wrote], but it’s yours now [sc. belongs to you].<sup>6</sup>

### **4.3 Compounded nouns**

On the unspecific account, the determinate meaning does not supply specific information about how the satisfiers of the nouns are related. If we know that a carpet cleaner cleans carpets whereas a vacuum cleaner cleans using a (partial) vacuum, or that traffic lights control rather than illuminate traffic, or that a band saw does not saw bands but saws with a band, this is not semantic knowledge, but non-semantic knowledge, concerning cleaners, traffic and saws, knowledge of which of the ways for an unspecific content to be true is most likely to be in question. No doubt there are or could be traffic lights which illuminate traffic, cleaning machines which use carpets as flails to clean, and saws designed specially for sawing bands.

In defence of this approach, consider a more complex case. A chemical purifier factory may be a factory that makes chemical purifiers, or one that uses chemical purifiers in making something else. Each of these two possibilities divides into two, for a chemical purifier may purify a chemical or purify by using a chemical. So there are (at least) four possibilities for chemical purifier factories. On a covert indexical view of compounded nouns, a context in which “A chemical purifier factory is to be built at the end of your garden” is intelligible is one which selects just one out of (a minimum of) four possibilities, and understanding the utterance involves knowing which was selected. This seems a highly implausible view. Once you learn that a chemical purifier factory is to be built at the end of your garden, you know enough to start making a fuss and writing to your MP without needing to know exactly what form the horror will take. Unspecific meaning seems just the right account.

Suppose that context somehow makes salient the pair of relations <using, making>, so that, according to a covert indexicality view, an utterance of “A chemical purifier factory is to be built at the end of your garden” in that context is true iff a factory which makes things which are used to purify a

6 There is also a shift from type to token.

chemical is to be built at the end of your garden. Suppose that in fact what is to be built at the end of your garden is a factory which makes things which purify by using a chemical. It seems to me nonetheless true that a chemical purifier factory is to be built at the end of your garden; this is inconsistent with a covert indexical account.

#### 4.4 Verbs

There probably is a transitive/non-transitive ambiguity in many verbs, including “runs” and “smokes”. Within non-transitive running, there are many kinds, as illustrated in 4(a). Many will typically be hard to access in a context. Perhaps some ways are excluded by restrictions relating to semantic categories. But we should not be too swift to think we know the exclusions. A recent cost-reducing innovation in the undertaking business involves chemically liquefying the corpses and running them into the public drains. John, once dead and in the grip of the innovators, can run like warm butter. There is no need to recognize semantic (as opposed to business) innovation here, or any change of meaning.

4(b) illustrates transitive/non-transitive ambiguity in “smokes”, and so does not as such pose a threat to the compositionalist. Within the transitive use it is surprising how unspecific the meaning turns out to be, as the following puzzle illustrates: how can a cigarette be half-smoked without being shorter than before? To smoke a cigarette you normally put one end in your mouth, light the other, and puff. To smoke a salmon, you normally soak it in brine for one and a half hours, hang it in your chimney, and when it is dry light little piles of oak chips at the bottom of the chimney and leave it for eight hours. There are different ways of smoking, but no ambiguity in the word “smoke”. You *could* smoke a cigarette by soaking it in brine for one and a half hours, hanging it in your chimney, and when it is dry lighting little piles of oak chips at the bottom of the chimney. Then half-smoking a cigarette, smoking it for a mere four hours, may not shrink it much or at all. This way of smoking cigarettes is much better for your health than the normal one. Trying to smoke a salmon in the way cigarettes are usually smoked is unlikely to be successful; this shows that one understands what it would be to try this, and reinforces the claim that there is no relevant ambiguity in “smoke”.

#### 4.5 The Travis effect

This can suggest some radical conclusions about the nature of language, of which failure of compositionality is at the conservative end of the spectrum. I believe that at least some of these cases can be dealt with by seeing meanings as suitably unspecific, though other cases are best described in terms of covert indexicality (and perhaps yet other cases are

genuine counterexamples to compositionality). I start by elaborating the examples a little.<sup>7</sup>

5(a) “Those leaves are green.” Brown leaves have been painted green. Suppose we are part of a commission inspecting Vietnam to determine whether Pentagon denials that it has used defoliants are true. Brown leaves are signs of the early stages of the action of defoliants. Travis suggests that, in this context, the truth about the painted leaves is that they are not green. The sentence has a meaning such that the facts just described make it false. Now suppose that we are trying to select camouflage material. Only green things will do, and more or less anything green will do. In this context, Travis suggests, the truth about the very same leaves is that they are green. The sentence has a meaning such that the same leaf-related facts make it false. Yet the sentence is not ambiguous, so it has only one meaning, it relates only to the leaves and the leaves are in the same state in both circumstances. This seems like a contradiction. Travis avoids it by denying that there is a proper conception of meaning which determines the truth conditions of what is said on an occasion (even allowing for the kind of context sensitivity manifested by pronouns).

I suggest that “Those leaves are green” is true in both cases, but that in the first a participant who came to learn that it is true would jump to the conclusion that it is made true in the normal way, rather than the exceptional way. This participant would be led astray; but one can easily be led astray by the truth (as by Desdemona’s handkerchief). The meaning of “green” is unspecific: there must be a green surface, but the meaning is indifferent to how deep the colour runs and how the surface got to be that colour. We generally make normal assumptions about these things, just as we assume that “John smokes” is true in virtue of John smoking cigarettes or cigars or a pipe, rather than in virtue of frequently smoking salmon. But we have no difficulty in seeing that these normal assumptions may fail to hold.

5(b) “The ball is round.” A first time spectator at a squash game asks if the ball is round. He wants to know whether squash resembles soccer, in being played with a round ball, or rugger, in being played with a ball which is not round. The right answer to the question is “Yes”, even if the ball is currently far from round thanks to having been hit against the wall.

In a contrasting case, a manufacturer of squash balls is trying out a new material. For the trials, tiny transmitters have been inserted into the skin of the ball to measure reactions to deformations. The instrument adjacent to the court, which is supposed to register the signals, is flat, which is as it should be if the ball is not at that moment deformed. The technician asks if the ball is

7 Sources are Travis (1985, 1994, 1996). Travis (1994) in effect denies compositionality for utterances or thoughts, on the grounds that many pragmatic determinations operate differently from those at work in determining reference for indexicals.

round. If the ball is at that point ovoid through being against the wall, the right answer is “No”.

In these cases, I find it hard not to accept Travis’s judgements of truth value, and this marks a difference from the other cases we have considered. For example, it was claimed that “Italian book” is ambiguous; but it was not explicitly claimed that “That’s an Italian book” could be both true and false of the same book, depending on which meaning is selected in the context. Had this claim been made, I would have rejected it.<sup>8</sup> But in the present case I find myself compelled to agree that “The ball is round” is made both true and false by the same ball in the same condition, depending on the context. This rules out the account in terms of unspecific meaning, for within the classical perspective which I accept, but which Travis is trying to undermine, an utterance with a single meaning, however unspecific, cannot be made to have opposite truth values by the same facts.

This leaves three options: either this is a failure of compositionality, or there is some lexical or structural ambiguity we have not yet considered, or there is a hidden contextual variable taking different values in the different cases. I offer a sketchy development of the last option.

Perhaps the present tense introduces a covert indexical for a stretch of time. An utterance containing “is  $\phi$ ” has as logical form “is  $\phi$  for at least  $t$ ” where the variable over temporal intervals is contextually determined (or determined in part by context and in part by the semantic character of  $\phi$ ). Polar values for this variable are *at this very moment*, and *in a general way*. John is writhing in the dentist’s chair. Is he happy? Not right now; but he is in a general way. Each of these values is probably also vague.

The two squash ball utterances pick up different values for  $t$  as a function in part of the concerns and interests of the participants. Hence the interpreted utterances say different things; and so, by one standard, differ in meaning. There is no counterexample to compositionality, any more than there is (or so we are supposing for the sake of the present discussion) in the fact that “Elle a faim” can be used to say different things.

## 5 Conclusion

While it can hardly be doubted that some measure of compositionality of understanding of utterances obtains, there is no apriori guarantee that it is universal. Even the staunchest defenders of compositionality admit this in their recognition of “idiom”: you cannot derive the standard colloquial meaning of “kicked the bucket” from a proper specification of the meaning

<sup>8</sup> One can hear an utterance of these words as both true and false if one supplies different implicit qualifications, but such a case is merely like “Clinton is good (as a statesman) and not good (as a husband)”.

of its parts. So we are left with a somewhat messy question: how far do these failures extend? In this essay I have considered some reasonably common *kinds* of apparent failure: locutions widespread enough that if compositionality failed for them, the claimed general compositionality for English would require serious qualification.

The counterexamples have taken the form of expressions which allegedly have different meanings or readings, even though there is no relevant lexical or structural ambiguity among their parts. In most cases, there are very many ways in which utterances containing the expressions could be true, and sometimes there seems no way of containing or listing all possibilities in advance. This tends to reduce the attractiveness of attempts to find some hidden lexical or structural ambiguity, for lexical and structural meanings must be fixed and known in advance. Approaches in terms of covert indexicality provide an appropriate kind of flexibility, but they suffer the defect that they predict that an utterance cannot be understood unless a specific value is assigned to the hidden variable, and in many cases this seems at variance with the facts. The claim that meaning is often unspecific deals well with a number of cases, though it cannot handle ones in which we are convinced that the utterances of the same sentence with the same reference for all explicit elements can have opposite truth values, as a function of context. In such cases, the most plausible option for the compositionalist, unless ambiguity can be detected in the components, is a covert indexicality view.

The issue of the extent to which our language is compositional is not to be decided by general apriori reasoning, but by detailed examination of specific cases, and careful attention to the methodology of various possible descriptions of them.<sup>9</sup>

9 Thanks to Ruth Kempson for discussion of these issues and to the Leverhulme Trust for a Senior Research Fellowship during the tenure of which this essay was written.

# Sense without reference

---

## I Introduction

Many people think that Frege allowed that expressions could have sense yet lack reference. The question I wish to raise is how one could justify the claim that a systematic description of natural language will make essential use of such a view.

The project might be thought doubly unsuitable for a conference on Frege. First, Frege himself probably had rather little interest in systematic accounts of the semantics of natural languages. However, I take it to be acceptable to see whether a thinker's ideas can be applied even to areas for which they were not originally designed. Second, there is some interpretative doubt about whether Frege held that expressions genuinely having sense, expressions capable of contributing to the expression of genuine thoughts, could lack reference, or at any rate, whether he held it for long. I am inclined to the view that he did hold it when he wrote "On Sense and Reference", where he says of the definite description, "the least rapidly convergent series", that it "has a sense but demonstrably lacks a reference". However, by the time of the piece called "Logic", dated 1897 and published only posthumously, he says that a sentence containing an expression lacking reference expresses at best a mock thought, a "Scheingedanke", and this would seem to be something which is not a thought.<sup>1</sup> So I think that there is a case to be made for saying that at least by that time he had abandoned the view that sense without reference was possible. I will not engage in this exegetical issue,\* relying on the clear statement in "On Sense and Reference" to justify labelling "Fregean" the doctrine that sense is possible without reference.

1 Despite Bell's contrary opinion (Bell 1990), the soundings I have taken among native speakers suggest that a Schein-*F* is something intended to seem an *F* even though it isn't an *F*. Thanks to Max Kölbel for discussion. Diamond (1984) offers a very useful discussion of Frege's position.

\* [Added note] See "Introduction: Departing from Frege", pp. 9–14, above.



This paper argues for an affirmative answer to its question as applied to proper names, as this expression is commonly used nowadays: semantically simple singular terms.<sup>2</sup> I argue that one can justify using a description of natural language which is “Fregean” in just the following respect: it makes essential use of the possibility of empty proper names. In deference to Russell’s view that “what does not name anything is not a name”, I shall label any denial of this “Russellian”. A full account would have at least three parts: (1) an attack on arguments for the Russellian view;<sup>3</sup> (2) a semantic theory which gives a recursive specification of meanings or truth conditions in a way that does not discriminate between empty and non-empty names; and (3) an account of the notion of a name-using practice which, likewise, is neutral between the case in which the practice involves an empty name and the case in which it involves a non-empty one.

The second task can be accomplished in more than one way. One could adopt the kind of descriptivist theory commonly (though controversially) associated with the historical Frege. On this view, a singular term is or abbreviates some kind of descriptive, effectively qualitative, condition, and its use in “primary occurrence” in a truth requires the unique satisfaction of this condition. However, there is no need for the relevant condition to be qualitative, in the way taken for granted in “descriptivist” theories. A theory neutral on this point, and which I shall take as my model in the present development of a Fregean position, has been provided by Tyler Burge (1974). In the semantics he develops, names are treated by axioms like

$\forall x$  (“Hesperus” refers to  $x$  iff  $x = \text{Hesperus}$ ).

The setting is negative free logic: atoms with empty names are false, and universal and existential quantifier rules are modified. An axiom of the above form for an empty name like “Vulcan” is true, because the right hand side is false for each value of  $x$ , leading to the appropriate verdict that there is nothing to which “Vulcan” refers. One interesting feature of the theory, which I will carry through to my own discussion, is that the semantics associates

2 As Jonathan Barnes pointed out to me, this is a rather inadequate characterization of the relevant class of expressions. Most westerners have forename and family name, and it is hard to see how this complexity (if that is what it is) is to be characterized. Are book titles names of books? These issues deserve closer scrutiny. A related issue is that I exclude demonstratives. While I certainly think that it is important to consider how the considerations of this paper relate to the use of demonstratives, I am unsure that the issues are logically connected. Both the view that the class of singular terms fractures into names and demonstrative, and the view that both are subsumed under a single category of singular terms, appear consistent with the main claim of this paper, viz. that there is a unified category of (empty and non-empty) proper names. Thanks to François Récanati and Ian Rumfitt for discussion.

3 For which see essay IX above.

names neither with an object nor with a description (in the usual qualitative sense of “description”).<sup>4</sup>

Hostility to this kind of approach may come from a reluctance to consider alternatives to classical logic. The following is designed to disturb the reluctance. If one accepts a standard account of validity, according to which a valid argument is one such that, for each world at which the premises are true, so is the conclusion, one will reject the classical rule of universal instantiation, since, even though “Socrates” is (in fact) not empty, there are worlds at which “Everything is perishable” is true but “Socrates is perishable” is not (worlds at which Socrates does not exist).<sup>5</sup> Rejecting the classical rule of instantiation is at least a step towards a free logic.

By contrast to the Burge-style approach, Russellian orthodoxy starts with the idea that a name is to be associated with an object, so an axiom for “Hesperus” will be based on the idea that

the reference of “Hesperus” = Hesperus.

Since, in the usual versions of such theories, the metalanguage does not contain primitive function symbols or a referential description operator, a formalized version would look more like:

- (1)  $\exists x(\text{“Hesperus” refers to } x \text{ and } \forall y(\text{“Hesperus” refers to } y \rightarrow x=y) \text{ and } x = \text{Hesperus})$ .

4 Thus Bell (1990: 275), summarizing a sentence each from the *Grundgesetze* and the *Grundlagen* writes: “for Frege the sense of an expression is the condition that must be met by anything that is the reference of that expression. . . . there can be a fully determinate, coherent, and intelligible condition which . . . nothing fulfils”.

The envisaged semantics also undermines any ultimate difference between reference and satisfaction. What we call reference is just satisfaction under a condition (like being Hesperus) capable of being satisfied by at most one thing.

Suppose the kind of haecceities invoked by the Burge-style theory I envisage turned out to be reducible to qualitative properties. Should one conclude that there is no difference between a Burge-style theory and a descriptive theory? The conclusion does not follow. I see a semantic theory as designed to state things which if known by speakers would explain their behaviour. But one might know something of the form “this has property F” without knowing anything of the form “this has property G” even if the property F is the property G. The space between a conventional descriptivist position and a Russellian one disappears only if conditions like being Hesperus really abbreviate qualitative conditions (as perhaps Russell himself thought). Thanks to David Sosa and Eric Loomis for discussion of this issue.

5 The argument was mentioned to me (though not endorsed) by Yannis Stephanou. The case is only *prima facie*. Classical instantiation could be defended in a number of ways, including (as David Wiggins suggested) shifting to binary quantification, which would require the premise that Socrates is a thing; or, more simply and classically, by guiding one’s formalization of the premise by the formulation “for all  $x$ , if  $x$  is a thing, then  $x$  is perishable”.

I call such axioms “bearer-specifying” because they affirm that the name has a bearer and go on to identify it. They cannot truly be affirmed by the theorist if the object language name has no bearer. In the next section, I give some general considerations in favour of a Fregean semantics, a semantics which can supply true axioms for empty names.

## 2 Some general Fregean considerations

Suppose one believes that one can get close to specifying the knowledge involved in understanding a sentence in terms of knowledge of what it would be for the sentence to be true.<sup>6</sup> What would be the corresponding thing to say about the knowledge involved in understanding a referring expression, say a name? A natural answer is that it is to know what it would be for the name to refer. This is not knowledge of what the name refers to, but rather knowledge of conditions under which it would refer, knowledge, that is, of how something would have to be in order to be what the name refers to. Just as one might make some progress in saying what it would be to understand “snow is white” in terms of knowing that the sentence is true iff snow is white, so one might make some progress in saying what it would be to understand “Hesperus” in terms of knowing that the name refers to something iff that thing is Hesperus. In short, if the meaning of a sentence is its truth condition, a cognate thought is that the meaning of a name is its reference condition; and this analogy is one general consideration in favour of a Fregean approach.

Suppose, however, that we start with the Russellian idea that a semantic theory should associate a name with an object. Running the analogy in the other direction would lead to the absurd suggestion that semantic theory should associate a sentence with a truth value. Current orthodoxy inexplicably associates sentences with conditions but names with objects. The only plausible way to restore the analogy is to associate names also with conditions rather than objects.

A semantic theorist, as radical interpreter, must immerse himself in the language-using practices of his subjects. Vague as the notion of immersion may be, it is natural to suppose that some degree of it is sufficient for understanding. In that case, nothing else is necessary. So unless immersion in a practice covertly requires the existence of a referent if the practice involves a name, the Fregean view is imposed by the data of interpretation. The main

6 I take it for granted that what it would be for “Hesperus is visible” to be true is not the same as what it would be for “Phosphorus is visible” to be true. This approach is, to put it optimistically, non-reductive: the distinction in question cannot be extracted merely from the nature of truth but must come from some “intensional” notion, in my opinion, following for example McDowell 1977 or Davies 1981a, that of propositional attitudes.

part of this paper is devoted to trying to establish that it is not the case that the notion of a name-using practice requires names to have bearers.

We can give examples of how easily empty names can be introduced, whether as fiction, jest or through error, and point out that we feel a need to teach our children the “correct use” of various actual empty names (“Vulcan”, “Santa Claus”), which implies that they are meaningful. We can insist that there is parity at the level of explaining behaviour. We can explain why many adults are excited by the thought of a trip to *Paris* in terms of their expectations that Paris is beautiful and has excellent restaurants, where the evidence for the relevant beliefs derives in part from the subjects’ “Paris”-utterances. In just the same way, there is a *prima facie* case for saying that we can explain why children are excited at Christmas in terms of their expectations that *Santa* will bring them presents, where the evidence for the relevant beliefs derives in part from the children’s “Santa”-utterances.

Cases in which the population under study is agnostic about, for example, whether there ever was such a person as Homer, or is divided on the question, are particularly striking. The Russellian theorist would need to resolve the issue. From his perspective, if “Homer” is empty, semantic theory has nothing to say about it: the activities relating to this mere sound do not constitute a name-using practice, and ordinary sentences containing it lack truth conditions. Yet it seems clear that semantic theory should be able to describe the relevant behaviour without risking falsification by the eventual discovery that the sceptics were right and there really is no such person as Homer.

We could not expect a semantic theorist to explore the historical origin of every name on the University’s register to see if it is genuine or is, rather, like “Paul R. Zwier” (Larson and Segal 1995: 161); nor need he be an astronomer, which he would have to be to distinguish “Neptune” from “Vulcan”; nor a theologian, which he would have to be to determine which, if any, of his subjects’ names for gods are empty; nor a chemist, which he would have to be to distinguish “phlogiston” and “ether” from “heat” and “air”; nor a literary theorist with sound views on the authorship of *Odyssey* and *Iliad*. To suppose otherwise is not merely implausible but potentially incoherent, for on a natural view the relevant investigations would take for granted that the names are intelligible, and would be guided by what that meaning is. This would underwrite the possibility of intelligible questions *whether* there is such a person as Paul R. Zwier, *where* Vulcan is supposed to be, if it exists, *what* phlogiston is meant to be like, and *who* Homer was, if anyone. The questions to be investigated are naturally expressed in a way that makes essential use of, and thus presupposes the intelligibility of, the names in question. Semantic theory is one thing, specialist knowledge of non-semantic fact another.

A preference for a Fregean over a Russellian theory can be motivated for non-empty names. It would seem possible that the users of such a name, say “*a*”, which in fact refers to *a*, should fail to know that *a* exists, even if they have true beliefs to this effect. The failure of knowledge might derive from a

deviant link in some causal chain, or from a serious lack of confidence (one can select an explanation to fit one's theory of knowledge). A Russellian semantic theorist is required to make an explicit affirmation of the existence of *a*. Since the theorist should affirm only what he knows, he is required to have knowledge that outstrips that of the speakers whose knowledge he is trying to describe. Moreover, in attributing to the speakers implicit knowledge of the semantic theory, he is attributing to them knowledge which, by hypothesis, they lack. No such contradiction threatens Fregean theories.

One can view a semantic axiom as if it were a stipulation governing the use of an expression. One cannot stipulate things into existence, so a Russellian axiom affirming the existence of an entity to which a name refers could not count as an axiom of semantic theory. By contrast, one can stipulate an "at most one" condition, as a Fregean theory does (for example, as developed by Burge). For one can stipulate that a tie for victory is defeat: if there is a group containing more than one candidate for meeting an "at most" condition, like that of being Hesperus, and each candidate in this group is as good as any other in the group but is a better candidate than any of the candidates outside the group, then no candidate counts as meeting the condition.

The familiarity of many of these arguments has not made them persuasive,<sup>7</sup> especially to Russellians. I think that more persuasive considerations emerge from a look at the nature of name-using practices.

### 3 Name-using practices

A necessary condition for the adequacy of an account of language is that it be able to provide an adequate description of its use, and this means in particular that provision should be made for a distinction between the practice of using the name "Aristotle" for the philosopher and the practice of using it for the tycoon. This distinction will either figure explicitly in semantic axioms for names, or will at least supply guidance about how the theorist is to reach and understand such axioms. If there is no sense without reference, one can individuate by appealing to the referent, just as I did in setting up the problem. But once one grants sense without reference, this cannot in general be the right way to individuate (though its correctness in the cases in which it is correct ought to follow from a correct general condition).

The general answer I offer is that name-using practices are individuated by their source in acts of name-introduction, where such acts can associate the

7 These considerations all have a *prima facie* character, for if there is a direct and decisive argument for the claim that a name must name, then this conclusion must somehow be accommodated in the methodology of semantic theorizing. Hence arguments against arguments for the view that every intelligible name must have a bearer are complementary to those offered here, and are offered in essay IX above.

new name with at most one object. This last condition ensures as a consequence that practices of using non-empty names with different bearers are different practices (though the converse does not hold), and this meets the requirement at the end of the previous paragraph. This source-based approach has two main rivals: a referent-based approach, characteristic of Russellian and Kripkean theories; and an information-based approach, characteristic of description theories of names. I suspect that it is dissatisfaction with theories of the latter kind that have driven people towards object-based, Russellian or Kripkean, accounts of name-using practices. What this paper supplies is the sketch of an alternative to both informational and object-based accounts.

Spelling this out in full detail is quite hard, but, in defence of the Fregean approach, I shall suggest that the task is not made significantly easier by having referents in the story. Kripke (1972) sketched a picture of name-using practices which start with a baptism, and are propagated causally from that starting point. What I shall suggest is that we can arrive at a Fregean source-based account by using essentially the Kripkean picture, but without thinking of the originating events, the baptisms, as essentially relating to objects.

### 3.1 A simple model

A first rough shot at stating Kripke's picture might be:

Two acts of using the name-type  $N$  belong to the same name-using practice iff there is an object  $x$  and a causal transmission relation  $R$  such that both acts are related to  $x$  by the ancestral of  $R$ .\*

The structure of his proposal is that there is an initiating event and a transmission relation. I say I can adjust the account of the initiating event and help myself to the same transmission relation.

For a Kripkean the transmission relation serves to put the recipient in epistemic contact with the referent, and this might make it seem that I cannot make use of the same transmission relation. But this is not so: this feature of the transmission relation results from the two Kripkean claims that transmission transmits knowledge of meaning, and that knowledge of the meaning of a name involves epistemic contact with its referent. It is only thanks to the second of these views that the transmission relation transmits knowledge of the referent. Once this is deleted, as it must be for a Fregean view, the transmission relation is neutral.

\* [Added note] This formulation would be better if quantification over the transmission relation took widest scope: there is a causal transmission relation,  $R$ , such that two acts of using the name-type  $N$  belong to the same name-using practice iff there is an object to which both are  $R$ -related.

A Kripkean picture of an early part of a practice might be as follows:



Figure 1

The filled circle represents the baptism, the square represents the object, and the open circles represent two subsequent uses belonging to the practice, in virtue of being related to the baptism by the transmission relation.<sup>8</sup> Subsequent uses are related not only to the baptism but also to the baptised object. However, it is unclear that the relation to the object could do any work: so long as the baptism does invest the name with a meaning, we need only check whether subsequent uses are related appropriately to it in order to determine that they belong to the same name-using practice. The Kripkean referent thus appears to be idle in the account of the unity of a name-using practice, provided that distinct baptisms can be distinguished independently of which object, if any, is baptized.

The alternative I propose for a Fregean account adopts Kripkean ideas about transmission, while eliminating the object. This means that a “baptism” may be a baptism of nothing, which is verbally awkward but just registers the Fregean view that a name can be intelligibly introduced even if it names nothing.

Justifying the choice of this approach would require ruling out approaches based upon associated information, as would be implied by description theories of names. My reasons for ruling out such theories are of familiar kinds, which I will not rehearse directly here, though §3.2 and §3.3 reveal some inadequacies with description theories.

### **3.2 Duplicated sources: qualitatively the same information may inform different practices**

Even when name-using practices coincide in what information is invoked, they constitute different practices if they originate in different objects. Suppose there are two speakers, S1 and S2, each of whom knows just one of the twins, Jim and Tim, and each calls the one he knows Harry. S1’s “Harry”-related information derives exclusively from Jim, S2’s from Tim. Qualitatively,

<sup>8</sup> The transmission relation is indicated by a black line. The distinct relation which holds between a baptiser and the baptised object is shown by a grey line.

their information may be identical (both affirm “Harry is tall”, “Harry is happy” etc.). But if they encounter one another, their apparently harmonious use of “Harry”-sentences provides merely an illusion of understanding: because their information derives from different sources, they are using the name in different ways (or, semantically, there are two different names with the same spelling). This result does not depend upon difference of object, for there is also difference of source: the “Harry”-uses of S1 and S2 lack a common source. The irrelevance of the object is made plain in the diagram (Figure 2).



Figure 2

One can imagine a structurally similar case for empty names. Two name-introducing rumours might coincide in information, but if they have different sources, I think we should rule that two names have been introduced and that there is a mere illusion of understanding when the distinct practices meet.

### 3.3 Information involved in distinguishing practices?

Must we not appeal to associated information in distinguishing one who has authentically joined a practice from one who has tried but failed? Does not one who sincerely affirms “London is the capital of France” give one reason to suspect he does not understand “London”? Likewise, does not one who sincerely affirms “Santa Claus is a planet which affects the orbit of Mercury” give one reason to believe he does not understand “Santa Claus”?

In both cases, this is merely defeasible evidence of lack of mastery. In the one case it would be defeated by, for example, the discovery that while in other respects a normal user of “London”, the speaker had been carried away by a dream about a new phase of British colonization. In the other case, it would be defeated by, for example, the discovery that while in other respects a normal user of “Santa Claus”, the speaker had come to believe an Ovidian fantasy in which the sledge driver had offended the gods by excessive jingling, and in punishment had been transformed into a silent planet.

It may seem that although in the case of a non-empty name there is no information whose possession is required for understanding, this is not so for empty names. Precisely because there is no bearer, contact with which could be involved in what makes for understanding, surely there must be a body of



information which plays a crucial role. This idea may be encouraged by the accidental fact that some empty names are associated with very little information (“Vulcan”), and others, though suffering no dearth of information, generally permit access to the full richness only to those who have passed through a narrow gateway of information (“Shylock”). These are not essential features of empty names. We could well imagine that the speculation about Vulcan was not rapidly quashed, but continued over several generations, the information being enriched by various myths, so that some later users are quite ignorant of the basis of the original postulation; so some competent users might rationally doubt that Vulcan was a planet. Likewise, it could be the case that a competent user of “Shylock” should be unaware of the Shakespearean origin, and suppose the name to refer to some nineteenth century miser. He would, of course, have needed to learn the name from a competent user, for example, a user well aware of the Shakespearean origin; but he might never have known, or have once known but forgotten, that Shylock was a fictional character. The hypothesis that he has just started up a new and non-Shakespearean fictitious use of a like-sounding name would be discredited by his having some recognizable Shakespearean line on such matters as who Shylock’s daughter was and the kinds of contract into which he entered. One such line would be simple belief, but another would be recognition that others hold such beliefs, together with an account of their falsehood. (“That pound of flesh stuff was obviously just an anti-Semitic exaggeration of a normal, if exigent, business arrangement.”)

Here is a story about how a name-introducing rumour might begin, which reveals the way in which information may diverge, and follows the simple model. An over-imaginative, or self-deceiving, or evil tongue,  $T_0$ , may start a rumour which is embellished by others. The rumour is that there’s a dragon, Fiamma, who lives in the mountain just south of the village and whose preferred diet is human babies. You hear the rumour from  $T_1$  and I hear it from  $T_2$ , each of whom heard it, on separate occasions, from  $T_0$ . You say that Fiamma is green, trusting to  $T_1$ ’s embellishment, and I say she is red, trusting to  $T_2$ ’s. By some standard, our Fiamma-related information has different origins; but there is a standard which rules that these different bodies of information have the same source, in  $T_0$ . This is the standard we need: it correctly represents us as disagreeing about Fiamma’s colour. It rules as it does because, although the information that Fiamma is green is new to  $T_1$ , and the information that Fiamma is red is new to  $T_2$ , both these pieces of information were intended by their producers to link to Fiamma,<sup>9</sup> so the “ultimate” origin lies further back, with  $T_0$ . This is the source that is invoked by my proposal.

9 The context of this occurrence of “Fiamma” is intensional: one cannot infer the existence of Fiamma from the existence of this intention. Our need to use names we know to be non-referring in describing intentions and other attitudes makes a strong case for the Fregean approach.

The view that no piece of information need be shared by all users in a practice has its analogue for empty names. Those who think that one can understand “Russell” without even knowing that its bearer is human will think that one can understand “Fiamma” without even knowing that its bearer is supposed to be a dragon: perhaps the rumour will develop, so that some say Fiamma is a gorgon, some say she prefers adults to babies, some that “she” is really a he, and some elders, wishing to reduce panic, claim he is vegetarian. These people disagree among themselves on the facts while agreeing on language, provided that their use of the name “Fiamma” has a common origin, and that each intends to speak of what the others speak of.

### 3.4 Fusion

Fusion of name-using practices provides a case in which an object of reference arguably dominates source in individuation.

Suppose that one group of speakers (a) sees a mountain from the north side and calls it “Everest” and another group (b) sees it from the south side and calls it by a name which, coincidentally, sounds and is spelled the same. At first, the groups do not meet. But then the route through the range is discovered. North-siders and south-siders talk to each other freely using “Everest”. There is a strong intuition that what began as distinct name-using practices will eventually fuse. One way to describe this would be to say that at first there were two practices and then there was one. The object theory can account for this, as there is a single object; the source theory apparently cannot, for there are distinct sources.<sup>10</sup>

A similar structure can arise in the case of empty names, with a diagram just like Figure 3, save with the object omitted. Coincidentally, among both those who live on the north side and those who live on the south side of the mountain a rumour springs up about a dragon, Fiamma, who lives in the mountain above the village and whose preferred diet is human babies. At first, the groups do not meet. But then the route through the range is discovered.

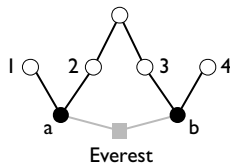


Figure 3

10 Another version, pointed out to me by Maite Ezcurdia, involves two scientists postulating the “same” object as unobserved cause of some phenomenon: distinct sources but, eventually, a single practice.

North-siders and south-siders talk to each other freely using “Fiamma”. There is a strong intuition that what began as distinct name-using practices will eventually fuse. The object theory cannot account for any of this, as it cannot allow that there are any name-using practices in the story. But the story also presents problems for the source theory, as there are two sources but, on one account of fusion, ultimately a single practice.

I believe that in both cases there is only an appearance of understanding in the initial encounters. For example, if (Figure 3) person 2 from the (a) practice meets person 3 from the (b) practice on the ridge, as contact is first established, it may seem as if they understand one another’s “Everest”-utterances. If I am right to hold that this is only an illusion of understanding, the source theory at least gets the early history right.

Communication can transmit knowledge. Suppose person 2 expresses knowledge with the words “Avalanche yesterday Everest”. Imagine that the first two words are not intelligible to 3, but that 2 explains them by signs and dumb-show. In connection with the third word, however, 3 indicates that he has no need for any explanation. Suppose that, before the utterance, 3 did not know that there was an avalanche on Everest yesterday. Suppose also, for *reductio*, that 3 understands what 2 said. Then, on any adequate account of knowledge by testimony, he ends up knowing that there was an avalanche on the north side of Everest on the previous day. However, it was just an accident that the sound 2 used for Everest coincided in its contribution to truth conditions with the similar sound as used by 3. Things could easily have been otherwise: the north-siders might easily have used “Everest” for the next mountain along. The method implicitly adopted by 3, homophonic translation, was not reliable. Knowledge is not acquired by unreliable methods. So 3 did not end up knowing that there had been an avalanche on Everest the previous day. The explanation is that 3 did not know what 2 said.

The source-based account thus gets the early part of the story right: at the beginning, there is at best an illusion of understanding. To do justice to this, the object-based account has to add some further condition. It is not enough for understanding that speaker and hearer derive their use of a name from a common object: in the present case, at least, it seems that they need to know that there is an object to which they both intend to refer. The object-based account thus needs some further condition, not contained merely in causal history, but most naturally to be found in the intentions, beliefs or knowledge of the speakers, in order to get the first part of the fusion story right; in order, that is, to allow that merely deriving the use of a name-type from a common object is not enough for the uses to belong to a single practice. The connection is that uses which belong to a single practice permit immediate and genuine understanding.

The difficulty for source-based accounts is to explain the fusion of distinct practices. A constraint upon a successful account is that it should respect the intuition that nothing more is needed for (a)-siders and (b)-siders to under-

stand each other, and thus for their practices to fuse, than for them to know that they both use “Everest” for the same mountain. There are various options, among which I suggest we take seriously an analogue of what it is natural for object-based theories to say in order to explain the distinctness of the practices before fusion: in addition to the origin of uses we must also look to the intentions and beliefs by which they are governed. The most obvious suggestion for object-based theories is that there should be a single object about which participants in an exchange belonging to the same practice aim to speak. The analogue for source-based theories suitable for a Fregean is that participants in such an exchange must intend to speak of a single object. The idea is that the shift in the scope of the quantification over objects means that although, setting aside fiction, the users must believe there is something of which they both speak, the theorist describing this use need not.

This seems an entirely proper thing to say about the empty case. There is no object such that either community can know that “Fiamma” stands for it. On the other hand, we do have some inclination to represent the fused situation as one in which both north- and south-siders intend to speak of the same dragon. Theorists can make sense of this without committing themselves to the existence of dragons. A sign of the presence of these coincident intentions is that north-siders view south-siders as well as north-siders as belonging to the community with whose practice they wish to accord and as potential suppliers of Fiamma-related information; they want to talk about whatever all these people, south-siders included, want to talk about. For south-siders, the position is analogous. In the absence of a dragon, the intentions cannot be realized. But they can be sustained by that fact that users have no evidence, or at least no decisive evidence, that there is more than one dragon (or less than one).

Returning to the “Everest” case: if, on the ridge, one party utters “Everest” while pointing to the mountain, and the other expresses agreement, mutual understanding is assured. On the present view, the explanation is that they come to appreciate that in using this word they are trying to speak of the same object. More happens: they come to know that there is an object concerning which both parties are trying to speak when using the word “Everest”. But this additional knowledge, though relevant to whether the parties are so related that they can transmit non-semantic knowledge, is, I claim, not required for semantic knowledge. For the latter, it is enough to appreciate the coincidence of the intentions.

What can “coincidence of intention” amount to, if not to one of the following: both intentions involve the same object, or both involve the same content? The first option is unavailable to the Fregean theorist, for the coincidence must be possible even in the absence of an object. The second option threatens to be circular, if the coincidence in content is in part determined by an agreement about how a name is to be used. There is a third option. In the “Fiamma” case, each party thinks: what the other is trying to speak of in

using “Fiamma” is just what I am trying to speak of when I use that name. This commits the speakers, but not the theorist, to belief in coincidence of object. Just this belief is what seems crucial in the “Everest” case. The difference is that in one case the belief is false, whereas in the other it constitutes knowledge. Given that a Fregean view is committed to typical serious uses of empty names being ones which involve false belief, it would not be surprising if false beliefs are among those that bind users together into a common name-using practice. These practices resemble other social practices in this respect: it is enough for the relevant people to believe that something is so for it to be so. In the present case, it is enough, in normal circumstances, for people to believe they are party to a common practice for this to be so.

Some of the problems for the source-based account arose from the supposition that after fusion has occurred there is just a single practice, rather than two fused practices. The very notion of *fusion* needs careful handling, since it can lead to contradiction. Suppose distinct things,  $x$  and  $y$ , fuse to  $z$  at point  $p$ . We are tempted to say that to fuse is not to cease to exist, so that  $x$  and  $y$  continue to exist beyond  $p$ . But then it may seem that both must be  $z$ , and we may be driven to the contradiction that  $x \neq y$  and  $x = z$  and  $y = z$ . There are two coherent descriptions: what we call  $z$  is really two things, though occupying a region into which just one thing of its kind will normally fit; or to fuse is to cease to exist. To apply this to a real example: Woodstock Road meets Banbury Road at St Giles, which carries on to Carfax. One option is to say that there are two roads which occupy a common space through St Giles: St Giles is both (a short stretch of) Banbury Road and (a short stretch of) Woodstock Road. Another option is to say that there are three roads: fusing is ceasing to exist. There seems to me no reason to think that every case of fusion must be treated in the same way. We are free to see how best to understand the fusion of name-using practices, and thus check on the consistency of the final picture with a source-based account.

On the first option, the fusion of the name-using practices means that there are two practices occupying the same “region”, geographical and linguistic. Typical users in the post-fusion phase are masters of both, and their utterances are contributions to both. This is straightforwardly consistent with there being two sources. There is some oddness in speaking of two practices after fusion, when everything goes so smoothly; and some oddness in saying that a single speech-act involving a single occurrence of a name involves two (specially related) uses; but then there is some oddness in speaking of two roads occupying the same space. The oddness derives from this option about how to understand fusions, rather than from the source view as such. It would require one to say that the previously discussed coincidence of intentions is only necessary and not sufficient for a single practice, since such intentions can be common to two fused practices; indeed, such intentions partially constitute the fusing force.

On the second option, we should say that there are three name-using

practices in north-side/south-side cases. The idea that a new practice is inaugurated by the meeting on the ridge is not wholly implausible. But it is very implausible that the previous practices cease to exist, and that stay-at-home south-siders, those who had had no contact, even indirect, with those who had encountered north-siders, have unwittingly embarked on a new practice. The old uses are adequately sustained by those who were not party to the historic meeting.<sup>11</sup> The best development of this position seems to be to imagine the three practices as overlapping for a while in space and time, with the newest practice gradually driving out both the old ones. If this option seems best, the source-based account can adapt to it by seeing the inauguration of the new practice as precisely that: a new source, from which subsequent uses derive. The gradual ousting of the old practices arises because best practice becomes that in which the participants aim to speak of whatever the wider community, their side as well as ours, aims to speak of.

In sum, any difficulties here spring from the notion of fusion; they are not special difficulties for the source-based mode of individuating name-using practices.

### 3.5 Confusions

A given utterance may be beholden, in ways normally suitable for making it part of a name-using practice, to two sources. In some of these multiple source cases, one may be at a loss (or in dispute) about how to individuate name-using practices. I learn the name “Harry” in the presence of Jim, but, without my realizing the shift, it’s Tim I mostly meet thereafter. My original “Harry”-related information comes from Jim; but the majority of my “Harry”-related information comes from Tim. To whom do I refer when I use “Harry”? The options are: Jim, Tim, both, neither, there’s no fact of the matter. To rephrase: should one count my possibly idiosyncratic name-using practice as one whose name refers to Jim, to Tim, to a Jim–Tim fusion, or to nothing? Or should one say there is no fact of the matter?

If Figure 4 is appropriate to these cases, it may seem that object-based individuation is at no advantage compared to a source-based approach.

As there are as many sources as objects, it may seem that this case will be equally problematic for both approaches. However, the object-theorist is at an advantage when it comes to explaining the distinction between this case (*confusion*) and fusion. The top part of Figure 4 corresponds to the top part of Figure 3, yet for Figure 3 I assumed that all theorists would wish to say that

11 The phenomenon of deference could cause changes in a practice by, as it were, “action at a distance”. If an expert to whom others defer changes his practice, their practice is automatically changed without their having any direct causal connection with the new usage. I can consistently stipulate that the example under discussion is not like this.

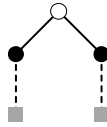


Figure 4

after fusion we had a fully coherent situation of mutual understanding: either a single practice, or two practices so fused as to resemble a single practice. Fusion can occur without confusion. It seems as if object-based theories can explain the difference: when two objects get into the story in this intermingled way, there is confusion; when practices relating to a single object mingle, there is fusion.

Fregean theorists will hold that confusion can occur even in the absence of an object, and so the object-based explanation just envisaged cannot be the full story. Suppose there are distinct and isolated “Fiamma”-using practices on each side of the mountain: they relate, as we feel inclined to say, to different dragons. Suppose a single intrepid trader finds a pass through the range and starts travelling back and forth. Should we say that his name-using practice belongs to the north-side practice? To the south-side practice? To both? To neither? Or that there is no fact of the matter? The same five options, and no reason of which I am aware to prefer one choice to another. So although I don’t have much to say about these cases, or an explanation of what makes them differ from fusions, it does seem to me that at least the most obvious object-based explanations will not work.

### 3.6 Fission

Evans gives what he calls a simplified version of the “Madagascar” example, which I shall treat as a degenerate case of the fission of name-using practices:

Two babies are born, and their mothers bestow names upon them. A nurse inadvertently switches them and the error is never discovered. It will henceforth undeniably be the case that the man universally known as “Jack” is so called because a woman dubbed some other baby with the name. (1973: 11)

We could diagram a slight generalization of this story as in Figure 5.

We envisage that some others (perhaps practical jokers on the nursing staff), represented by 1, continue to use “Jack” of the offspring, whereas the mother and various others, represented by 2, use it of the changeling, so that we have a case of fission. The mother’s later uses count as uses of the name

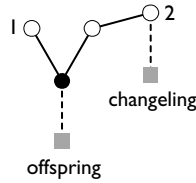


Figure 5

for the changeling, despite her link to her earlier practice. (In Evans's actual example, the first segment of the fission is not mentioned and perhaps does not occur: but we can see "bearer-shifting" as a degenerate case of fission.)

It indeed seems right to say that "Jack" refers to the man universally known as "Jack", in which case there is an apparent problem for a source-based account of the individuation of name-using practices: baptisms do not individuate name-using practices, for the baptismal use of "Jack" does not belong to the same practice as its subsequent use. After the switch, the mother started calling someone else's baby Jack, taking it to be her own. This use came to prevail. Unbeknownst to her, it inaugurated a new name-using practice, a distinct one from the short-lived practice inaugurated by a possibly more official baptismal ceremony. There are two names "Jack" in the story, just as there are (more than) two names "Aristotle".

Should a source-based theory count the mother's first use of "Jack" for the changeling baby as an unwitting baptism? This may seem to be inconsistent with the facts: she is surely making some kind of mistake when, on first confronting the little changeling, she utters "Jack is hungry", even if the child before her is hungry; she is certainly using the name intending to conform with her previous intentions; and she is in causal contact with her original baptism of her real child. As with fusion, it may seem more realistic to think of a gradual process, allowing that at first the mother is in error, but then at some indefinite later point, utterances of the kind just envisaged come to be true or false in virtue of how things are with the changeling rather than in virtue of how they are with her real offspring. It is not impossible to see how an object-based account might achieve this. But can a source-based account?

On her first encounter with the changeling, there is a radical failure in the mother's referential intentions. She is trying to conform her current use to her earlier use; but if an earlier use enabled her to refer to her child, a later use which does not enable this is hardly in conformity with that earlier use. We can account for the sense that the mother has made a mistake in terms of the failure of her intentions, rather than in terms of the failure of what the mother utters to be true (if the changeling is indeed hungry). What is gradual, I suggest, is not the inauguration of a new practice, but our appreciation of



this. We need to be convinced that this is not just a one-off error but a systematic departure; and we could know this only after some reasonable lapse of time, during which she still takes the changeling for her own offspring. If this does not happen, we do not judge the case to be the inauguration of a new practice, and indeed it would not have been: rather, it would have been simply a mistake, and subsequent uses would have conformed to the original practice. (Or if the offspring and the changeling were frequently swapped, we would be in the “confusion” cases already discussed.) Thus whether or not something counts as the inauguration of a new practice depends in part on what the future holds; just as whether or not a poisoning is a killing. When we think that the new practice gets a hold gradually, we are confusing the metaphysics of the situation (if this is indeed a usage which is destined to become stable, then it is an inauguration of a new practice) with what we can know about it (the relevant knowledge is available only later).

This approach is supported by a comparison with empty names.  $T_0$  started the rumour about the dragon Fiamma by saying “In the mountains to the south of the village, there lives a dragon, Fiamma”. Someone mishears this as “In the fountains near the mouth of the river, there lives a gorgon, Fiamma”. Under the misapprehension, she continues what she takes to be the same tale, elaborating on how the gorgon likes to startle the villagers who come to bathe. Inadvertently, she has started a new rumour, and a new name-using practice. The name “Fiamma” has divided, or, if the original dragon-rumour dies out, we can say that the name has, as it were, switched bearers. The explanation seems to be that although the second user is trying to conform her use to that of the first, in this she radically fails; and this unifies the explanation of why we count her as starting a new rumour (about a new beast) with the explanation of why we count Jack’s mother as having started a new name-using practice (about a new baby). In both cases, the event is an inauguration only if a practice descends from it; otherwise it is simply a failed attempt to continue an old practice.

#### **4 Source a disjunctive property?**

Since everyone agrees that, if a name is individuated purely syntactically, there are practices of using empty names and practices of using non-empty ones, the relevant notion of a source needs to mark out a non-disjunctive property of name-using practices. Yet I can imagine someone inspecting the concept of a source I have tried to develop and claiming that it is disjunctive.

I myself find it hard to engage with the objection, for I do not know what it is for a property (as opposed to a predicate) to be non-disjunctive in some absolute sense. What seems to me to matter is whether a property is treated disjunctively or not in some context. Thus the property of becoming a parent is treated disjunctively in English employment law: becoming a parent by becoming a mother is one thing, entitling the possessor to a decent period of

paid leave; becoming a parent by becoming a father is another, entitling the possessor to a much shorter period of leave. On the other hand, in another context, effectively the same property is treated non-disjunctively by the English welfare benefits system. Being a single parent entitles you to the same benefits whether you are a single parent by being a single mother or whether you are a single parent by being a single father. So the question I can understand is whether semantic theory treats the notion of source as disjunctive or not.

What I have recommended in this paper is that it be treated as non-disjunctive. Practices using empty and non-empty names are individuated in the same way, in terms of their source. Semantic axioms will take the same form for both cases. They will not be discriminated within the theory, whose form, therefore, is restricted by these considerations (e.g. non-Meinongian model theory cannot in any straightforward way do justice to empty names). While I recognize that my reasons for making this suggestion may fall short of being conclusive, I do not think that a reasonable response is merely to claim, as if reporting on the upshot of a metaphysical inspection, that the relevant notion of source is disjunctive. What would need to be done would be to give theoretical reasons for treating it as such.<sup>12</sup>

12 Thanks to the British Academy for contributing to my costs involved in travelling to the World Congress of Philosophy at Boston in 1998, where parts of this paper were presented; to audiences at the World Congress, at Philosophy Departments in the Universities of Lublin (the Catholic University), Texas (Austin), Mexico City (UNAM), Glasgow and St Andrews, and at the Frege Conference in Bonn in October 1998; and (in addition to those mentioned in other footnotes) to Stephen Barker, Herb Hochberg, Fraser McBride, Philip Percival, Sarah Sawyer, Yannis Stephanou, David Wiggins and Crispin Wright; and to Keith Hossack, who gave me very detailed and helpful comments on what would otherwise have been the final draft.

---

# Bibliography

---

- Beaney, M. (1996). *Frege: Making Sense*. London: Duckworth.
- Bell, D. (1990). "How 'Russellian' was Frege?". *Mind* **99**: 267–77.
- Bezuidenhout, A. (1997). "Pragmatics, semantic underdetermination and the referential/attributive distinction". *Mind* **106**: 1–34.
- Black, M. (1937). "Vagueness: an exercise in logical analysis". *Philosophy of Science* **4**: 427–55. Reprinted in his *Language and Philosophy*. Ithaca, NY: Cornell University Press, 1949: 25–58.
- Brandom, R. (1994). *Making It Explicit*. Cambridge, MA: Harvard University Press.
- Burge, T. (1974). "Truth and singular terms". *Noûs* **8**: 309–25. Reprinted in K. Lambert (ed.). *Philosophical Applications of Free Logic*. Oxford: Oxford University Press, 1991: 189–204.
- Campbell, J. (1987). "Is sense transparent?". *Proceedings of the Aristotelian Society* **88**: 273–92.
- Candlish, S. (1996). "Russell's theories of judgement". In R. Monk and A. Palmer (eds). *Russell and the Origins of Analytic Philosophy*. Bristol: Thoemmes Press, 103–35.
- Castañeda, H.-N. (1966). "'He': A study in the logic of self-consciousness". *Ratio* **8**: 130–57.
- Crimmins, M. and J. Perry (1989). "The prince and the phone booth". *Journal of Philosophy* **86**: 685–711.
- Davidson, D. (1967). "Truth and meaning". *Synthese* **7**: 304–23. Reprinted in his *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 1984: 17–36.
- Davidson, D. (1969). "On saying that". *Synthese* **19**: 130–46. Reprinted in his *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 1984: 93–108.
- Davidson, D. (1973). "Radical interpretation". *Dialectica* **27**: 313–28. Reprinted in his *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 1984: 125–39.
- Davidson, D. (1976). "Reply to Foster." In G. Evans and J. McDowell (eds). *Truth and Meaning*. Oxford: Oxford University Press, 33–41.
- Davidson, D. (1979). "Moods and performances". In A. Margalit (ed.). *Meaning and Use*. Dordrecht: Reidel, 9–20.
- Davidson, D. (1984). *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.
- Davies, M. (1981a). *Meaning, Quantification, Necessity: Themes in Philosophical Logic*. London: Routledge and Kegan Paul.
- Davies, M. (1981b). "Meaning, structure and understanding". *Synthese* **48**: 135–61.

- Diamond, C. (1984). "Frege against fuzz". In her *The Realistic Spirit: Wittgenstein, Philosophy and the Mind*. 1991: 145–77. Cambridge, MA: The MIT Press.
- Donnellan, K. (1972). "Proper names and identifying descriptions". In D. Davidson and G. Harman (eds). *Semantics of Natural Language*. Dordrecht: Reidel, 356–79.
- Dummett, M. (1973). *Frege: Philosophy of Language*. London: Duckworth. Page references to second edition (1981).
- Dummett, M. (1976). "What is a Theory of Meaning? (II)". In G. Evans and J. McDowell (eds). *Truth and Meaning*. Oxford: Oxford University Press.
- Dummett, M. (1983). "Existence". In D. P. Chattopadhyaya (ed.). *Humans, Meanings and Existences*. Delhi. Reprinted in Dummett 1993: 277–307.
- Dummett, M. (1993). *The Seas of Language*. Oxford: Oxford University Press.
- Evans, G. (1973). "The causal theory of names". *Proceedings of the Aristotelian Society: Supplementary Volume* 47: 187–208. Reprinted in Evans 1985: 1–24.
- Evans, G. (1975). "Identity and predication". *Journal of Philosophy* 72: 243–63. Reprinted in Evans 1985: 25–48.
- Evans, G. (1979). "Reference and contingency". *Monist* 62: 161–89. Reprinted in Evans 1985: 178–213.
- Evans, G. (1980). "Things without the mind". In Z. v. Straaten (ed.). *Philosophical Subjects: Essays Presented to P. F. Strawson*. Oxford: Clarendon Press. Reprinted in Evans 1985: 249–90.
- Evans, G. (1982). *The Varieties of Reference*. Oxford: Clarendon Press.
- Evans, G. (1985). *Collected Papers*. Oxford: Oxford University Press.
- Evans, G. and J. McDowell (eds) (1976). *Truth and Meaning*. Oxford: Clarendon Press.
- Ezcurdia, M. (1996). "Dynamic and coherent thoughts". *European Review of Philosophy* 2.
- Fine, K. (1975). "Vagueness, truth and logic". *Synthese* 30: 265–300.
- Fodor, J. and E. Lepore (2001). "Why compositionality won't go away: Reflections on Horwich's 'deflationary' theory". *Ratio* 14: 350–68.
- Foster, J. (1976). "Meaning and truth theory". In G. Evans and J. McDowell (eds). *Truth and Meaning*. Oxford: Oxford University Press: 1–32.
- Frege, G. (1884). *Die Grundlagen der Arithmetik*. Breslau.
- Frege, G. (1892a). "On sense and meaning".\* Reprinted in B. McGuinness (ed.). *Collected Papers on Mathematics, Logic and Philosophy*. Oxford: Basil Blackwell, 1984: 157–77.
- Frege, G. (1892b). "On concept and object". *Vierteljahrsschrift für wissenschaftliche Philosophie* 16: 192–205. Reprinted in B. McGuinness (ed.). *Collected Papers on Mathematics, Logic and Philosophy*. Oxford: Basil Blackwell, 1984: 182–94.
- Frege, G. (1895). "A critical elucidation of some points in E. Schröder, *Vorlesungen über die Algebra der Logik*". *Archiv für systematische Philosophie* 1: 433–56.

\* This work is referred to as "On sense and reference" throughout this book. The German is "Über Sinn und Bedeutung". While "meaning" may be a good translation of many colloquial uses of "Bedeutung" it is inappropriate in this connection. Frege's question whether any sentence has a Bedeutung is very hard to understand when translated as the question whether any sentence has a meaning, but less hard to understand when translated as the question whether any sentence has a reference (or, even better, referent).

- Reprinted in B. McGuinness (ed.). *Collected Papers on Mathematics, Logic and Philosophy*. Oxford: Basil Blackwell, 1984: 210–28.
- Frege, G. (1896). Letter to Guiseppe Peano, 29 September 1896. In B. McGuinness and K. Kaal (eds). *Philosophical and Mathematical Correspondence*. Oxford: Basil Blackwell, 1980: 112–5.
- Frege, G. (1897). “Logic”. In H. Hermes, F. Kambartel, and F. Kaulbach (eds and trans.). *Frege: Posthumous Writings*. Oxford: Basil Blackwell, 1979: 126–51.
- Frege, G. (1906a). “A brief survey of my logical doctrines”. In H. Hermes, F. Kambartel, and Friedrich Kaulbach (eds and trans.). *Frege: Posthumous Writings*. Oxford: Basil Blackwell, 1979: 197–202.
- Frege, G. (1906b). “Introduction to Logic.” In H. Hermes, F. Kambartel, and F. Kaulbach (eds and trans.). *Frege: Posthumous Writings*. Oxford: Basil Blackwell, 1979: 185–96.
- Frege, G. (1906c). “17 Key sentences on Logic.” In H. Hermes, F. Kambartel, and F. Kaulbach (eds and trans.). *Frege: Posthumous Writings*. Oxford: Basil Blackwell, 1979: 174–5.
- Frege, G. (1914). “Logic in mathematics”. In H. Hermes, F. Kambartel, and F. Kaulbach (eds and trans.). *Frege: Posthumous Writings*. Oxford: Basil Blackwell, 1979: 203–52.
- Frege, G. (1918). “Logical investigations”. In B. McGuinness (ed.). *Collected Papers on Mathematics, Logic and Philosophy*. Oxford: Basil Blackwell, 1984: 351–72.
- Frege, G. (1923). “Compound thoughts.” In B. McGuinness (ed.). *Collected Papers on Mathematics, Logic and Philosophy*. Oxford: Basil Blackwell, 1984: 390–406.
- Frege, G. (1924/5). “Sources of knowledge of mathematics and the mathematical natural sciences”. In H. Hermes, F. Kambartel and F. Kaulbach (eds). *Frege: Posthumous Writings*. Oxford: Basil Blackwell, 1979: 267–77.
- Geach, P. T. (1957). *Mental Acts*. London: Routledge and Kegan Paul.
- Geach, P. T. (1962). *Reference and Generality*. Ithaca, NY: Cornell University Press.
- Goguen, J. A. (1969). “The logic of inexact concepts”. *Synthese* **19**: 325–73.
- Griffin, N. (1985). “Russell’s multiple relation theory of judgement”. *Philosophical Studies* **47**: 213–47.
- Griffin, N. (1986). “Wittgenstein’s criticism of Russell’s theory of judgement”. *Russell* **5.2**: 132–45.
- Griffin, N. (1993). “Terms, relations, complexes”. In A. D. Irvine and G. A. Wedeking (eds). *Russell and Analytic Philosophy*. Toronto: University of Toronto Press, 159–92.
- Harcourt, E. (1993). “Are hybrid proper names the solution to the completeness problem?”. *Mind* **102**: 301–13.
- Henry, D. P. (1984). *That Most Subtle Question*. Manchester: Manchester University Press.
- Higginbotham, J. (1994). “Priorities in the philosophy of thought”. *Supplementary Proceedings of the Aristotelian Society* **68**: 85–106.
- Horwich, P. (1998). *Meaning*. Oxford: Oxford University Press.
- Horwich, P. (2001). “Deflating compositionality”. *Ratio* **14**: 369–85.
- Hughes, C. (1996). “Giving the sceptic her due?”. *Epistemologia* **19**: 309–26.
- Janssen, T. M. V. (1997). “Compositionality”. In J. van Benthem and A. ter Meulen (eds). *Handbook of Logic and Language*. Amsterdam: Elsevier, 417–74.

- Kaplan, D. (1977). "Demonstratives". In J. Almog, J. Perry and H. Wettstein (eds). *Themes from Kaplan*. Oxford: Oxford University Press, 1989: 481–614.
- Kempson, R. M. and A. Cormack (1981). "Ambiguity and quantification". *Linguistics and Philosophy* 4: 259–309.
- Kripke, S. (1972 [1980]). *Naming and Necessity*. Oxford: Basil Blackwell, 1980. Lectures delivered in 1970 and originally published in D. Davidson and G. Harman (eds). *Semantics of Natural Languages*. Minneapolis: University of Minnesota Press, 1972.
- Kripke, S. (1973). John Locke lectures delivered at the University of Oxford. Unpublished typescript.
- Kripke, S. (1979). "A puzzle about belief". In A. Margalit (ed.). *Meaning and Use*. Dordrecht: Reidel, 239–83.
- Künne, W. (1992). "Hybrid proper names". *Mind* 101: 721–31.
- Larson, R. and G. Segal (1995). *Knowledge of Meaning: An Introduction to Semantic Theory*. Cambridge, MA: MIT Press.
- Lewis, D. (1972). "General Semantics". In D. Davidson and G. Harman (eds), *Semantics of Natural Languages*, Dordrecht: Reidel.
- Loar, B. (1988). "Social content and psychological content". In R. H. Grimm and D. D. Merrill (eds). *Contents of Thought*. Tucson: University of Arizona Press, 99–110.
- Locke, J. (1690). *Essay Concerning Human Understanding*. London.
- Mackie, J. L. (1973). *Truth, Probability and Paradox*. Oxford: Oxford University Press.
- Mates, B. (1950). "Synonymity". *University of California Publications in Philosophy* 25: 201–26. Reprinted in L. Linsky (ed.). *Semantics and Philosophy of Language*, Urbana, IL: University of Illinois Press.
- McDowell, J. (1976). "Truth conditions, bivalence and verificationism". In G. Evans and J. McDowell (eds). *Truth and Meaning*. Oxford: Clarendon Press.
- McDowell, J. (1977). "On the sense and reference of a proper name". *Mind* 86: 159–85.
- McTaggart, J. M. E. (1927). *The Nature of Existence*. Cambridge, Cambridge University Press.
- Millikan, R. (1991). "Perceptual content and Fregean myth". *Mind* 100: 439–59.
- Millikan, R. (1993a). "On mentalese orthography". In B. Dahlbom (ed.). *Demmett and his Critics*. Oxford: Blackwell, 97–123.
- Millikan, R. (1993b). *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.
- Millikan, R. (1993c). "White queen psychology". In Millikan 1993b: 279–363.
- Moltmann, F. (forthcoming). "Nominalizing quantifiers."
- Nozick, R. (1981). *Philosophical Explanations*. Oxford: Oxford University Press.
- Peacocke, C. (1976). "Truth definitions and actual languages". In G. Evans and J. McDowell (eds). *Truth and Meaning*. Oxford: Clarendon Press.
- Peacocke, C. (1998). "The modality of freedom". In A. O'Hear (ed.). *Current Issues in Philosophy of Mind*. Cambridge: Cambridge University Press.
- Perry, J. (1977). "Frege on demonstratives". *Philosophical Review* 86: 474–97. Reprinted in Perry 1993: 3–32.
- Perry, J. (1979). "The problem of the essential indexical". *Noûs* 13: 3–21. Reprinted in Perry 1993: 33–52.
- Perry, J. (1983). "Castañeda on *He* and *I*". In J. E. Tomberlin (ed.). *Agent, Language*

- and the Structure of the World: Essays Presented to Hector-Neri Castañeda and His Replies*, Indianapolis: Hackett Publishing Company. Reprinted in Perry 1993: 91–119.
- Perry, J. (1993). *The Problem of the Essential Indexical*. Oxford: Oxford University Press.
- Plantinga, A. (1997). “Warrant and accidentally true belief”. *Analysis* **57**: 140–5.
- Platts, M. (1979). *Ways of Meaning: An Introduction to Philosophy of Language*. London: Routledge and Kegan Paul.
- Ramsey, F. (1926). “Truth and probability”. Reprinted in D. H. Mellor (ed.) (1978). *Foundations*. 58–100. Cambridge: Cambridge University Press.
- Récanati, F. (1993). *Direct Reference*. Oxford: Blackwell Publishers.
- Récanati, F. (1995). “Quasi-singular propositions: the semantics of belief reports”. *Supplementary Proceedings of the Aristotelian Society* **69**: 175–93.
- Rosch, E. (1978). “Principles of categorization”. In E. Rosch and B. Lloyd (eds). *Cognition and Categorization*. New York: Wiley & Son, 27–48.
- Russell, B. (1903). *The Principles of Mathematics*. Cambridge: Cambridge University Press.
- Russell, B. (1910). “Some explanations in reply to Mr Bradley”. *Mind* **19**: 373–8. Reprinted in J. G. Slater (ed.) (1992). *The Collected Papers of Bertrand Russell*. London: Routledge, 354–8.
- Russell, Bertrand (1911a). “Le réalisme analytique.” *Bulletin de la société française de philosophie* **11**: 282–91. Reprinted in J. G. Slater (ed.) (1992). *The Collected Papers of Bertrand Russell*. London: Routledge, 410–32.
- Russell, B. (1911b). “Knowledge by acquaintance and knowledge by description”. *Proceedings of the Aristotelian Society* **11**: 108–28. Reprinted in his *Mysticism and Logic*. London: George Allen and Unwin, 1910, 1963: 152–67.
- Russell, B. (1912a). “On the relations of universals and particulars”. *Proceedings of the Aristotelian Society* **12**: 1–24.
- Russell, B. (1912b). *Problems of Philosophy*. Oxford: Oxford University Press.
- Russell, B. (1918–19). “Lectures on the Philosophy of Logical Atomism”. *Monist* **28**, 29. Reprinted in R. C. Marsh (ed.). *Bertrand Russell: Logic and Knowledge; Essays 1902–1950*. London: George Allen and Unwin, 177–281.
- Russell, B. (1919). *Introduction to Mathematical Philosophy*. London: George Allen and Unwin.
- Russell, B. (1959). *My Philosophical Development*. London: George Allen and Unwin.
- Russell, B. and Whitehead, A. N. (1910–13). *Principia Mathematica*. Cambridge: Cambridge University Press.
- Sainsbury, R. M. (1983). “On a Fregean argument”. *Analysis* **43**: 12–14.
- Sainsbury, R. M. (1995). “Vagueness, ignorance and margin for error”. *British Journal for Philosophy of Science* **46**: 589–601.
- Salmon, N. (1986). *Frege’s Puzzle*. Cambridge, MA: MIT Press.
- Sorabji, R. (1983). *Time, Creation and the Continuum*. London: Duckworth.
- Sosa, E. (1996). “Postscript to ‘Proper Functioning and Virtue Epistemology’”. In J. L. Kvanvig (ed.). *Warrant in Contemporary Epistemology: Essays in Honor of Plantinga’s Theory of Knowledge*. 271–80. Lanham MD and London: Rowman and Littlefield.
- Stalnaker, R. (1970). “Pragmatics”. *Synthese* **22**: 272–93. Reprinted in his *Context and Content*. Oxford: Oxford University Press, 1999: 31–46.

- Stanley, J. (2000). "Context and logical form". *Linguistics and Philosophy* **23**: 391–434.
- Strawson, P. (1957). "Propositions, concepts and logical truths". *Philosophical Quarterly* **7**. Reprinted in his *Logico-Linguistic Papers*. London: Methuen, 1971: 116–29.
- Strawson, P. (1959). *Individuals*. London: Methuen.
- Travis, C. (1985). "On what is strictly speaking true". *Canadian Journal of Philosophy* **15**: 187–229.
- Travis, C. (1994). "On constraints of generality". *Proceedings of the Aristotelian Society* **94**: 165–88.
- Travis, C. (1996). "Meaning's role in truth". *Mind* **105**: 451–66.
- Wallace, J. (1978). "Logical form, meaning, translation". In F. Guenther and G. Guenther-Reutter (eds). *Meaning and Translation*. London: Duckworth.
- Walton, K. L. (1978). "Fearing fictions". *Journal of Philosophy* **75**: 5–27.
- Walton, K. L. (1990). *Mimesis as Make-Believe*. Cambridge, MA: Harvard University Press.
- Wiggins, D. (1986). "On singling out an object determinately". In P. Pettit and J. McDowell (eds). *Subject, Thought and Context*. 169–80. Oxford: Clarendon Press.
- Wiggins, D. (1995). "The Kant–Frege–Russell view of existence: toward the rehabilitation of the second-level view". In W. Sinnott-Armstrong, D. Raffman and N. Asher (eds). *Modality, Morality and Belief. Essays in Honor of Ruth Barcan Marcus*. Cambridge: Cambridge University Press.
- Wiggins, D. (1999). "Names, fictional names and 'really'". *Proceedings of the Aristotelian Society: Supplementary Volume* **73**: 271–86.
- Williamson, T. (1994a). *Vagueness*. London: Routledge.
- Williamson, T. (1994b). "Definiteness and Knowability". *The Southern Journal of Philosophy* **33**: 171–91.
- Williamson, T. (1999). "On the structure of higher-order vagueness". *Mind* **108**: 127–43.
- Wittgenstein, L. (1921). *Tractatus Logico-Philosophicus*. London: Routledge and Kegan Paul.
- Wittgenstein, L. (1974). *Letters to Russell, Keynes and Moore*. Oxford: Basil Blackwell.





---

# Index

---

- Altham, Jimmy 26, 158  
ambiguity 107, 127, 138, 195–203  
anaphora 9, 136, 137–57, 178  
Anselm of Canterbury 159–60  
apriori 38, 97, 127–30, 132, 134  
assertion 12, 164, 185  
Ayer, A. J. 44
- Barker, Stephen 223  
Barnes, Jonathan 206  
Beaney, Michael 136  
Bedeutung 43, 128, 184, 225  
Bell, David 205, 207  
Bermúdez, José Luis 28  
Bezuidenhout, Anne 195  
borderline 16, 22–3, 71–84, 112, 120–1  
boundary *see* borderline  
Bradley, F. H. 103, 110  
Brandom, Robert 154  
British Academy 26, 180, 223  
Burge, Tyler 14, 162, 177, 206, 210
- Campbell, John 25, 125, 133, 134  
Candlish, Stuart 24, 103, 104–6  
Castañeda, Hector-Neri 146  
Chalmers, David 182  
class 71–2, 107  
communication 4–5, 24, 35, 46, 55–7, 90–4, 101, 163, 164–6, 216  
compositionality 6, 10–11, 19, 21, 27–8, 192–204  
concatenation 17, 108–10  
concept 13, 16–17, 59, 71–2, 76–8, 81–2, 112, 120–3, 152, 178–9; concept-word 178–9; without boundaries 22–3, 71–84  
Cormack, Annabel 196  
Crimmins, Mark 107, 194
- Daniels, Charles 91  
Davidson, Donald 6, 8, 17, 20–2, 30, 35–7, 39, 41, 80–1, 108–9, 185  
Davies, Martin 32, 42, 45, 66, 70, 152, 208  
definite description 44, 45, 48, 194, 195, 199, 205; and proper names 4–5, 10, 14, 23–4, 44, 85–99  
demonstrative: identification 47–9, 51–2, 165; pronoun 37, 45, 55, 69, 141, 148–9, 157, 193–4, 206  
Diamond, Cora 17, 205  
Donnellan, Keith 98, 195  
Dummett, Michael 4, 5, 7, 14–16, 17, 22, 30, 41, 44, 86, 155, 161, 173–4, 175, 184
- entities: meanings as 27, 107–10, 181–8, 192; senses as 3, 7, 27, 106, 158; vs things 190–1  
Evans, Gareth 5, 9–12, 14, 15, 22, 24, 27, 30, 35, 43–70, 95, 159–67, 169–77, 184, 195, 220–1  
existence 27, 47, 64–9, 95–7, 98, 164, 171–80, 210  
existential quantifier 148, 163, 179, 206  
extensional 29, 35–7, 43, 71–3, 80, 144, 145, 181–2, 184, 186, 187  
externalism 16, 19, 113, 121–4, 129–30, 132

- Ezcurdia, Maite 158, 215
- fiction 9–13, 64–9, 159, 167–74, 209, 214
- fission 220–1
- Fodor, Jerry 27, 192
- force 30, 31
- Foster, John 21, 35, 36–8, 41, 42
- Frege, Gottlob 1–19, 22, 23, 24, 25, 26, 27, 28, 35, 43–6, 47, 55, 56, 57, 62, 86, 87, 107, 125–36, 137, 155–8, 160–2, 169, 176, 183, 184, 192–3, 205–6
- function 16–17, 24, 75, 82–3, 107, 119, 191, 207
- fusion 215–20
- Geach, Peter 106, 145
- generality constraint 59
- Giaquinto, Marcus 70
- Goguen, J. A. 75
- Griffin, Nicholas 102, 103, 106, 109
- Guttenplan, Samuel 70
- Hale, Bob 158
- Harcourt, Edward 126
- Henry, Desmond Paul 160
- Higginbotham, James 142–3, 158
- Hochberg, Herb 223
- homophonic 5–7, 14, 18, 27, 80–1, 82, 83, 109–10, 138–9, 191, 216
- Horwich, Paul 27, 184–6, 189, 190, 191
- Hossack, Keith 124, 126, 136, 158, 223
- Hughes, Christopher 113, 116, 124, 158
- Hylton, Peter 91
- “I” 5, 9, 55, 56, 137, 139, 141, 146, 156, 194
- identification 45, 48–51, 53, 55–7, 148, 199
- identity 18, 43, 96, 97–8, 128–32, 134–5
- indexical 2, 5, 9, 23, 24, 26, 126, 137–9, 143, 146–53, 157, 193–5, 196, 197, 202; covert indexicality 193–5, 197–8, 199–204
- indirect speech 7–8, 139, 184
- information 45, 48, 61, 62, 92, 96, 99, 100, 131–2, 143, 164–7, 199, 212–15; -invoking 47–55, 65, 160, 161–3, 169, 171, 187
- interpretation 6, 20, 28, 40, 52, 142–4, 169, 196, 198, 208; radical 6, 20–1, 30, 35, 39, 41
- interpretive (truth theory) 20, 21, 22
- Irvine, Andrew 23, 101
- Jackson, Frank 152
- Janssen, T. M. J. 192
- Jenkins, Gary 70
- Kamp, Hans 42, 70
- Kaplan, David 9, 107, 128, 194
- Kempson, Ruth 196, 204
- Klein, Peter 163
- knowledge 1–2, 3, 8, 10, 15–16, 18–19, 20, 21, 25, 27, 30–1, 33–4, 38, 39, 40, 42, 54, 60, 61, 88, 90, 92, 95, 112–24, 127–8, 131–4, 137, 150, 154–5, 160, 162–3, 164, 174, 181, 184, 187, 188, 189, 200, 208, 209, 210, 211, 216, 217–18
- knowledge\* 181–4, 186–7
- Kölbel, Max 205
- Kripke, Saul 3, 18, 23, 28, 57–8, 67–8, 85–7, 93–4, 95, 96, 98–100, 134, 160, 211–12
- Künne, Wolfgang 126
- Ladimeji, O. A. 158
- Larson, Richard 209
- Lepore, Ernest 27, 192
- Leverhulme Trust 28, 204
- Lewis, David 41
- Loar, Brian 94, 180
- Locke, John 83
- Lockwood, Michael 101
- logic 13, 14, 21, 73, 75–6, 80, 81, 133–4, 207; free 14, 27, 162, 170, 177, 206, 207
- logical constant 5, 110
- logical form 8, 26, 36, 66, 68, 91, 109, 132, 133–4, 139, 141, 145, 146, 171, 178, 181, 183, 194, 197, 203
- Loomis, Eric 207
- McBride, Fraser 223

- McDowell, John 6–7, 14, 30, 35, 36, 39, 40, 41, 45, 49, 57, 70, 133, 184, 208
- McFetridge, Ian 70
- McGinn, Colin 70
- Mackie, John 106
- McTaggart, J. M. E. 152–4
- make-believe *see* pretence
- Martin, Michael 136, 158, 180
- Mates, Benson 7, 8
- meaning 1, 6, 18, 19–21, 24, 25, 27, 28, 29–42, 43, 45, 46, 47, 71, 81, 86, 87, 88–9, 90, 94, 96, 102, 104–11, 112, 121–2, 127–30, 137–9, 141, 143–4, 147–8, 155–7, 160, 164, 171, 181–91, 192, 193–204, 206, 208, 209, 211, 212
- meaning\* 184–6
- meaning rationalism 127–30
- meaningless 45, 95–6, 128, 172
- Mill, J. S. 19
- Millikan, Ruth 25–6, 125, 127–33, 136
- mode of identification *see* identification
- mode of presentation *see* presentation, mode of
- Moltmann, Friederike 185
- Morton, Adam 158
- MRTJ (the multiple relation theory of judgement) 102–3, 106, 107, 109, 110
- name *see* proper name
- Neale, Stephen 180
- Nozick, Robert 113, 115, 124
- O’Hear, Anthony 70
- paratactic 8, 36, 37, 185
- PDP (parallel distributed processing) 82
- Peacocke, Christopher 33–4, 124, 136
- Percival, Philip 158, 223
- Perry, John 56, 138, 151–2, 157, 194
- photograph model 22, 57–62
- Plantinga, Alvin 114
- Platts, Mark 35
- presentation, mode of 2–5, 10, 26, 43–4, 127–8, 133, 155–8
- pretence 11, 27, 54–5, 64–9, 159, 168–75
- proper name 3–5, 9–16, 18–19, 23–4, 28, 35, 44, 46–7, 55–6, 58, 65, 85–101, 105, 108, 145, 152, 160–79, 186–7, 189, 206–23; descriptive 45; empty 9–14, 19, 27, 28, 44, 96, 160, 164, 166–8, 170, 172, 176–9, 206–23; fictional 65, 67–8, 159, 167–72, 174, 177; logically proper 45, 93; mock 11–12; -using practice 16, 18, 28, 58, 100, 164, 166–8, 171, 209–23
- proposition *see* unity of the proposition
- propositional attitude 1, 7, 8, 10, 40, 138, 185, 208
- prototype 81–2
- quantifier 45, 48, 50, 69, 91, 94, 148, 153, 162, 179, 190, 196, 206, 207
- quasi-: fear 167–9; sense 68–9; understanding 65, 68, 167–70, 174, 177
- Quine, Willard van Ormond 27, 35, 41, 190
- radical interpretation *see* interpretation
- reading 28, 195, 196–9, 204
- “really” 27, 66–9, 159, 171–80
- Recanati, François 107, 206
- recognitional abilities 30–2, 34
- reference 1–4, 6–14, 16–18, 22–3, 26, 27, 35, 43–6, 52, 53, 55, 57–8, 86–7, 89, 90, 92, 94, 96, 98–9, 100, 101, 108, 125–6, 128, 130, 131, 133–6, 137, 139, 140, 145–9, 152, 155, 157, 162, 164, 165, 169, 182, 183, 202, 205, 207–8, 210, 215
- referent 1–4, 6–7, 9–19, 24, 25, 27, 28, 43–7, 52–5, 57–8, 92–3, 94, 96, 100, 101, 161, 163–4, 167, 173, 183, 192, 192, 199, 208, 210–12, 225
- reliability conditional 25, 112–15, 124
- Rosch, Eleanor 82
- Rumfitt, Ian 206
- Russell, Bertrand 4–5, 15, 17, 22, 23–4, 44–7, 57, 63, 69, 85–101, 102–11, 148, 160, 181, 206, 207
- Russellian 14, 24, 45–8, 50–3, 55, 57, 61–2, 63–5, 159–60, 167–8, 172, 174, 206–11
- Russell’s principle 47–8, 58–9, 61, 160

- Salmon, Nathan 2  
samesaying 8, 10, 18, 26, 139, 140  
saturation 14, 17, 24, 107  
Savile, Anthony 70  
Sawyer, Sarah 223  
Schröder, E. 13  
Searle, John 44, 86  
Segal, Gabriel 158, 180, 209  
semantic: theory 1–3, 5–11, 14, 16, 19, 28, 29–42, 43, 44, 71–2, 76, 78, 79–81, 83, 95–6, 98, 105, 108, 113, 114, 123, 125, 126, 127–30, 132, 133, 137, 139, 140–5, 155, 156, 162, 180, 184, 193, 194, 196, 200, 201, 203, 205–6, 207, 208–10, 217, 223; structure 30–5, 41; value 43–4, 159  
sense 1–19, 23, 25–6, 28, 40, 41, 42, 43–4, 46, 47, 55, 57, 68, 87, 125–36, 152, 155–8, 161, 183–4, 192, 193, 198, 205, 207; customary and indirect 7–8, 155, 184; senses as entities 3, 27, 28, 126, 158, 183, 192; without a referent 2, 9–14, 27, 169, 205–23  
set *see* class  
sharp 14, 16, 22, 71–5, 77, 80, 83, 112, 120  
singular term 1, 2, 10, 11, 19, 27, 43–55, 57, 64–5, 67–8, 134, 144, 145, 160–4, 169, 172, 178, 185, 206  
Smiley, Timothy 26  
Sorabji, Richard 150, 158  
Sosa, David 207  
Sosa, Ernest 25, 113, 114, 117, 124  
source 18, 28, 50, 92, 161, 165, 210–23  
Stalnaker, Robert 195  
Stanley, Jason 194  
Stephanou, Yannis 180, 207, 223  
Stoneham, Tom 136  
Strawson, Peter 25, 72, 86, 125, 133, 150  
Sullivan, Peter 180  
Sutton, Jonathan 180  
Textor, Mark 158  
thought, mock (Scheingedanke) 11–12, 205  
tracking 113–15, 124  
Travis, Charles 16, 180, 196, 201–3  
truth condition 24, 25, 41–2, 63–4, 66–9, 80, 87, 93–4, 96–7, 98–101, 107–10, 142, 152, 154, 159–60, 169–70, 171, 174–6, 191, 194–5, 202, 206, 208, 209, 216  
truth theory (T-theory) 6, 17, 20, 26, 30, 37, 40–2, 141–2, 144–5  
twin earth 64  
type/token 8–9, 18, 34–5, 125–6, 133–6, 137, 139, 141, 143, 145–9, 151, 152, 156–7, 193, 200  
unity of the proposition 17, 24, 102–11  
vagueness 16, 22–3, 72–83, 112, 121–2  
verb, twofold nature of 102–3, 105, 109  
verificationism 15, 40  
Wallace, John 35  
Walton, Kendal 167–8, 171  
Wedeking, Gary 23, 94  
Wiggins, David 26, 27, 42, 158, 171–2, 177–80, 207, 223  
Williamson, Timothy 19, 25, 74, 112–13, 119–20, 121, 122, 124, 158  
Wittgenstein, Ludwig 27, 103, 105–6, 182, 188, 191, 193  
Wright, Crispin 223  
Zalta, Ed 183